


## RESEARCH ARTICLE

# Bayesian measurement-error-driven hidden Markov regression model for calibrating the effect of covariates on multistate outcomes: Application to androgenetic alopecia

Amy Ming-Fang Yen<sup>1</sup> | Hsiu-Hsi Chen<sup>2</sup> 

<sup>1</sup>School of Oral Hygiene, College of Oral Medicine, Taipei Medical University, Taipei, Taiwan

<sup>2</sup>Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan

**Correspondence**

Professor Hsiu-Hsi Chen, Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Room 533, No 17, Hsueh Road, Taipei 100, Taiwan.  
Email: chenlin@ntu.edu.tw

**Funding information**

Ministry of Science and Technology, Grant/Award Numbers: MOST 106-2118-M-532-001-MY2, MOST 106-2118-M-038-002 -MY2, MOST 106-2118-M-002-006-MY2 and MOST 107-3017-F-002-003; Ministry of Education, Grant/Award Number: NTU-107L9003

Multistate Markov regression models used for quantifying the effect size of state-specific covariates pertaining to the dynamics of multistate outcomes have gained popularity. However, the measurements of multistate outcome are prone to the errors of classification, particularly when a population-based survey/research is involved with proxy measurements of outcome due to cost consideration. Such a misclassification may affect the effect size of relevant covariates such as odds ratio used in the field of epidemiology. We proposed a Bayesian measurement-error-driven hidden Markov regression model for calibrating these biased estimates with and without a 2-stage validation design. A simulation algorithm was developed to assess various scenarios of underestimation and overestimation given nondifferential misclassification (independent of covariates) and differential misclassification (dependent on covariates). We applied our proposed method to the community-based survey of androgenetic alopecia and found that the effect size of the majority of covariate was inflated after calibration regardless of which type of misclassification. Our proposed Bayesian measurement-error-driven hidden Markov regression model is practicable and effective in calibrating the effects of covariates on multistate outcome, but the prior distribution on measurement errors accrued from 2-stage validation design is strongly recommended.

**KEYWORDS**

Bayesian, calibration, hidden Markov model, Markov regression model, measurement error

## 1 | INTRODUCTION

Markov regression models have been extensively used in identifying the effects of significant state-specific covariates on transitions between multiple states of disease process with specific applications including cancer,<sup>1,2</sup> diabetes mellitus,<sup>3</sup> Alzheimer's disease,<sup>4</sup> and stroke.<sup>5</sup> More importantly, elucidating state-specific covariates accounting for the dynamics of multistate outcome plays an important role in recently proposed concept of precision medicine for prevention, surveillance, treatment, and therapy of disease in question,<sup>6,7</sup> which is equivalent to the concept of N-of-1 trials for individually tailored therapy.<sup>8</sup>

The copyright line for this article was changed on 24 May 2018 after original online publication.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2018 The Authors. *Statistics in Medicine* published by John Wiley & Sons Ltd.

While those models are very useful, one of complications is that the classification of multistate outcomes may be prone to measurement errors, particularly when the multistate outcomes are defined by imperfect diagnostic tools in population-based researches. Such measurement errors may lead to the biased effect size of covariate in association with step-by-step multistate transitions, which have been scarcely addressed before 2000.

It is noteworthy that these measurement errors can be viewed as the discrepancy between observed state and hidden (true) state under the context of hidden Markov process. The classical application of hidden Markov model (HMM) to handling the misclassification of true states in previous studies have recourse to the mixed HMM.<sup>9-12</sup> Zhang and Berhane<sup>9</sup> developed a Bayesian-mixed HMM to calibrate the effect of covariates on each state transition on the basis of the observed outcomes of asthma measured by self-reported questionnaire with multilevel data structure by using the generalized model underpinning that consists of 3 parts: prevalence of asthma, transition probability between true states, and the misclassification probability. The results demonstrated that the use of mixed HMM is more accurate than the traditional logistic regression model. Dunson<sup>10</sup> and Miglioretti<sup>11</sup> proposed the latent transition models coupled with the generalized linear framework to deal with multiple outcome measurements, which could be a mixture of count, categorical, and continuous observed response, given the latent transitions of true states. Altman<sup>12</sup> applied such a kind of mixed HMMs to capture the overdispersed and autocorrelated count data resulting from extra-Poisson process given the unobserved true disease status. All these approaches to estimating error probabilities are based on the mixed HMM coupled with the framework of generalized linear model whereby the relevant covariates can be incorporated.

To estimate the parameters of the mixed HMM, Bayesian Markov chain Monte Carlo (MCMC) method has been widely used over the past 2 decades. However, the so-called label-switching problem due to the symmetry of likelihood function has been encountered when it comes to Bayesian MCMC method because posterior distribution of the mixed HMM is invariant when the label-switching method is used. To solve this problem, several methods have been proposed, including the artificial identifiability constraints of the parameters on mixture distribution or those on component, relabeling algorithms coupled with decision theoretic approach with Bayesian MCMC or expectation-maximization algorithm, and probabilistic relabeling algorithm.<sup>13,14</sup> It could be argued that such an issue still remains when estimating component-specific parameters and interpreting marginal posterior densities are of great interest. Moreover, over-parameterization would not be avoided and the label-switching problem would become complicated if the number of hidden state increases due to biological property.

Furthermore, while the mixed HMM is applied to dealing with measurement errors, interests still center on how measurement errors affect regression coefficients of relevant covariates but not on the evolution of dynamic disease process. Chen et al<sup>15</sup> and Jackson et al<sup>16</sup> studies applied the HMM to the disease natural history of cancer with the consideration of measurement errors resulting from inaccuracy of population-based screening data. However, both of HMMs were targeted to the transition of natural history, but their models were not focused on how these measurement errors affect the effects of covariates on the multistate outcomes of disease natural history.

Instead of using the mixed HMM, the alternative proposed method here is the use of forward-backward or Viterbi algorithm of HMM to treat the probability of measurement error as the emitting probability and the transition probability as the probability of transition between true states. Doing so enables one to model the transition of disease progression in continuous-time Markov process and also to calibrate the effect size of relevant covariate making allowance for measurement errors. The most advantage of using forward-backward algorithm or Viterbi algorithm is that it enables one to estimate the most possible dynamic latent true health state and the sequence of health state for any individual.

To be commensurate with the expedient use of Bayesian approach but not to resort to using artificial identifiability constraints and determined or probabilistic relabeling algorithm,<sup>17</sup> an alternative approach is to use a 2-stage validation design with a small sample of the pilot study to calibrate these errors.<sup>18</sup> This further justifies the use of the Bayesian approach to incorporate prior information obtained from the pilot validation study into data on the main study to calibrate the estimates of regression coefficients with adjustment for these measurement errors. In the language of epidemiology, the biased effect resulting from misclassification is often quantified by differential (covariate dependent) or nondifferential misclassification (covariate independent) to reflect the direction of effect size before and after correcting these measurement errors. However, little is known about how the differential and nondifferential misclassification affects the multistate disease process.

The remaining section of this paper is organized as follows. Section 2 delineates the multistate measurement-error-driven hidden Markov regression model for the effect of state-specific covariates on the multistep disease progression. We developed a new recursive relationship of conditional probabilities between observed state, true state, and measurement error with hidden Markov underpinning given repeated observed data. Section 3 shows a 2-stage validation design with Bayesian underpinning. The informative prior on the measurement errors of multistate outcomes were derived in the first stage and were combined with the empirical data that forms the likelihood to build up the posterior distribution

in the second stage. The Bayesian MCMC simulation with Gibbs sampling scheme was applied so as to estimate the parameters for statistical inference. Section 4 presents how these measurements errors affect the effect of covariates of interest on the transitions of multiple states given differential and nondifferential misclassifications and provide the simulation algorithm and the results for assessing the magnitudes of underestimation and overestimation given various scenarios of misclassification. Section 5 illustrates the application of the proposed Bayesian measurement-error-driven hidden Markov regression model to data from a community-based survey on androgenetic alopecia (AGA) with a 2-stage validation design. We also specified how the information of classification errors of AGA from the first stage with gold standard involved has been incorporated into the stochastic process built for the large-scale community-based survey in the second stage of main study. Finally, Section 6 gives the discussion on methodological thoughts, the biases caused by differential and nondifferential misclassification in 3-state HMM, and the empirical findings of AGA example.

## 2 | MODEL SPECIFICATION

### 2.1 | Multistate Markov model for k-state progressive disease

Suppose we have a  $k$ -state Markov model for the progression of disease from state 1 to state  $k$  with state space  $\Omega = \{1, 2, \dots, k\}$  realized by the random variable  $Y(t)$  to represent the state occupied at time  $t$ . Such a progressive model is a phase-type distribution, which assumes the disease is progressive from state 1 to state  $k$ . To estimate a series of successive transitions, discrete-time or continuous-time Markov model can be considered.<sup>19,20</sup>

With the continuous-time Markov process, following the notation of Kalbfleisch and Lawless,<sup>21</sup> the instantaneous transition rates were specified between states. The transition from state  $i$  to state  $j$  is denoted by  $q_{ij}$  with the following definition:

$$q_{ij}(t) = \lim_{\Delta t \rightarrow 0} \frac{P(Y(t + \Delta t) = j | Y(t) = i)}{\Delta t} \quad i = 1, 2, \dots, k, \text{ and } j = 1, 2, \dots, k, \quad (1)$$

which is the  $i$ th row and  $j$ th column element in the instantaneous transition  $k \times k$  matrix  $\mathbf{Q}$ . It should be noted that the  $i$ th diagonal element  $q_{ii}(t) = -\sum_{j=1}^k q_{ij}(t)$ . An absorbing state has a zero row vector in  $\mathbf{Q}$  matrix. Accordingly, the transition probability matrix, say  $\mathbf{P}$ , can be derived as

$$\mathbf{P}'(t) = \mathbf{P}(t)\mathbf{Q}, \quad (2)$$

subject to the initial condition,  $\mathbf{P}(0) = \mathbf{1}$ .

With a discrete-time Markov model, we need to describe the disease process with 1-step transition matrix,  $\mathbf{P}$ . Therefore, the transition probability in  $n$  step can be expressed as  $\mathbf{P}^n$ .

### 2.2 | Measurement-error-driven HMM

To consider the measurement errors between states, we introduce  $r_{ab}$  to denote the proportion of misclassifying state  $b$  (true state) as state  $a$  (observed state). We start with the first visit and decompose the observed probability of being classified as state 1,  $P(O_1^{(1)})$ , into  $k$  true disease status at the first visit,  $T_1^{(1)}, T_2^{(1)}, \dots$ , and  $T_k^{(1)}$ , representing state 1, 2, ..., and  $k$ , respectively, which enable us to estimate multiple types of measurement errors as described below. Accordingly, the first visit for the observed state 1 is decomposed into  $k$  components of conditional probabilities for those with true state 1 and correctly observed as state 1 ( $r_{11}$ ), denoted by  $P(O_1^{(1)}|T_1^{(1)})$ , for those with true state 2 but misclassified as state 1 ( $r_{12}$ ), denoted by  $P(O_1^{(1)}|T_2^{(1)})$ , ..., and for those with true state  $k$  but misclassified as state 1 ( $r_{1k}$ ), denoted by  $P(O_1^{(1)}|T_k^{(1)})$ .

In terms of HMM language,  $\gamma_{ab}$  is regarded as the emitting probability defined as

$$\gamma_{ab} = P(O_a|T_b) \quad (a, b = 1, 2, \dots, k). \quad (3)$$

The probability of having initial observed value  $O_a^{(1)}$  expressed by  $\gamma_{ab}$  is written as follows:

$$P(O_a^{(1)}) = \sum_{b=1}^k P(T_b^{(1)}) \cdot P(O_a^{(1)}|T_b^{(1)}) = \sum_{b=1}^k \pi_b(A) \cdot \gamma_{ab}. \quad (4)$$

where  $\pi_b(A)$  represents true prevalence of state  $b$  for those going to the first visit at  $A$  years of age.

To build up a recursive relationship between the observed probability at the second visit and that at the first visit with time interval  $t$ , the conditional probability of observing state 1 at the second visit includes  $\frac{k \cdot (k+1)}{2}$  possibilities from

1. true state 1 ( $T_1^{(1)}$ ) and observed as state 1 ( $O_1^{(1)}$ ) at first visit which is further divided into  $k$  scenarios following over time until the second visit; staying in state 1 until the second visit ( $T_1^{(2)}$ ), progressing to state 2 ( $T_2^{(2)}$ ) between the 2 visits but misclassified as state 1 ( $O_1^{(2)}$ ) at the second visit, ..., and progressing to state  $k$  between the 2 visits but misclassified as state 1 at the second visit and
2. true state 2 ( $T_2^{(1)}$ ) but misclassified as state 1 ( $O_1^{(1)}$ ) at first visit which is further divided into  $k-1$  scenarios given progressive assumption (ie, the regression to true state 1 is not possible); staying in state 2 ( $T_2^{(2)}$ ) between the 2 visits but misclassified as state 1 ( $O_1^{(2)}$ ) at the second visit, ..., and progressing to state  $k$  ( $T_k^{(2)}$ ) between the 2 visits but misclassified as state 1 ( $O_1^{(2)}$ ) at the second visit

$k-1$ . true state  $k-1$  ( $T_{k-1}^{(1)}$ ) but misclassified as state 1 ( $O_1^{(1)}$ ) at first visit which is further divided into 2 scenarios; staying in state  $k-1$  ( $T_{k-1}^{(2)}$ ) between the 2 visits but misclassified as state 1 ( $O_1^{(2)}$ ) at the second visit, and progressing to state  $k$  ( $T_k^{(2)}$ ) between the 2 visits but misclassified as state 1 ( $O_1^{(2)}$ ) at the second visit; and  $k$ .

True state  $k$  ( $T_k^{(1)}$ ) but misclassified as state 1 ( $O_1^{(1)}$ ) at first visit which stays in state  $k$  ( $T_k^{(2)}$ ) between the 2 visits but misclassified as state 1 ( $O_1^{(2)}$ ) at the second visit.

Therefore,

$$\begin{aligned} P(O_1^{(2)}|O_1^{(1)}) &= P(O_1^{(2)}, T_1^{(1)}|O_1^{(1)}) + P(O_1^{(2)}, T_2^{(1)}|O_1^{(1)}) + \dots + P(O_1^{(2)}, T_k^{(1)}|O_1^{(1)}) \\ &= \left\{ \left[ P(O_1^{(2)}, T_1^{(2)}, T_1^{(1)}, O_1^{(1)}) + \dots + P(O_1^{(2)}, T_k^{(2)}, T_1^{(1)}, O_1^{(1)}) \right] \right. \\ &\quad \left. + \left[ P(O_1^{(2)}, T_2^{(2)}, T_2^{(1)}, O_1^{(1)}) + \dots + P(O_1^{(2)}, T_k^{(2)}, T_2^{(1)}, O_1^{(1)}) \right] \right. \\ &\quad \left. + \dots + P(O_1^{(2)}, T_k^{(2)}, T_k^{(1)}, O_1^{(1)}) \right\} / P(O_1^{(1)}), \end{aligned} \quad (5)$$

which can be generalized as follows:

$$P(O_1^{(2)}|O_1^{(1)}) = \sum_{b=1}^k P(O_1^{(2)}, T_b^{(1)}|O_1^{(1)}) = \sum_{b=1}^k \sum_{l=b}^k \frac{P(O_1^{(2)}, T_l^{(2)}, T_b^{(1)}, O_1^{(1)})}{P(O_1^{(1)})}. \quad (6)$$

Note that the 2 summation expressions only yield  $\frac{k \cdot (k+1)}{2}$  components because some (like reversible transitions) are not admissible.

$$\begin{aligned} &\text{Since } P(O_1^{(2)}, T_l^{(2)}, T_b^{(1)}, O_1^{(1)}) \\ &= P(O_1^{(2)}|T_l^{(2)}, T_b^{(1)}, O_1^{(1)}) \cdot P(T_l^{(2)}|T_b^{(1)}, O_1^{(1)}) \cdot P(O_1^{(1)}|T_b^{(1)}) \cdot P(T_b^{(1)}) = P(O_1^{(2)}|T_l^{(2)}) \cdot P(T_l^{(2)}|T_b^{(1)}) \cdot P(O_1^{(1)}|T_b^{(1)}) \cdot P(T_b^{(1)}), \end{aligned}$$

which, following the definition of the emitting probability and the Equation (4), is expressed as

$$= \gamma_{1l} \cdot P_{bl}(t) \cdot \gamma_{1b} \cdot \pi_b(A) \quad (7)$$

$P(O_1^{(2)}|O_1^{(1)})$  becomes

$$\sum_{b=1}^k \sum_{l=b}^k \frac{\gamma_{1l} \cdot P_{bl}(t) \cdot \gamma_{1b} \cdot \pi_b(A)}{P(O_1^{(1)})}. \quad (8)$$

Extending the Equation (8), the general form for the probabilities of observing state  $v$  in the next time  $t_v$  given observing state  $u$  in the previous time  $t_u$  to accommodate irregular interval between 2 visits can be expressed as

$$P\left(O_v^{(t_v)} | O_u^{(t_u)}\right) = \sum_{b=1}^k \sum_{l=b}^k \frac{\gamma_{vl} \cdot P_{bl}(t_v - t_u) \cdot \gamma_{ub} \cdot \pi_b(A)}{P(O_u^{(t_u)})}. \quad (9)$$

The expression (9) is similar to the forward equation of HMM to derive the conditional prediction of state at  $t_v$  given the observed state at  $t_u$ . The similar argument can be expressed by Viterbi algorithm with the replacement of summation by maximum function as often stated in the conventional textbook of HMM.<sup>22</sup>

The above derivation can be decomposed into the detailed observations across true disease status for the first visit ( $\mathbf{U}$ ) and the transition between the  $(v-1)^{th}$  and the  $v^{th}$  visit ( $\mathbf{W}_v$ ) in matrix form as follows.

$$\mathbf{U} = \left( \mathbf{O}_1^{(1)} \quad \cdots \quad \mathbf{O}_k^{(1)} \right), \quad (10)$$

in which each element is a vector of length  $k$  for the composition of true disease states,  $\mathbf{O}_1^{(1)} = (\pi_1 \cdot \gamma_{11}, \dots, \pi_k \cdot \gamma_{1k})$ ,  $\mathbf{O}_2^{(1)} = (\pi_1 \cdot \gamma_{21}, \dots, \pi_k \cdot \gamma_{2k})$ , ..., and  $\mathbf{O}_k^{(1)} = (\pi_1 \cdot \gamma_{k1}, \dots, \pi_k \cdot \gamma_{kk})$ ,

$$\mathbf{W}_v = \begin{pmatrix} \mathbf{M}_{11} & \cdots & \mathbf{M}_{1k} \\ \vdots & \ddots & \vdots \\ \mathbf{M}_{k1} & \cdots & \mathbf{M}_{kk} \end{pmatrix}, \quad (11)$$

where the  $r$ th row and the  $c$ th column element,  $\mathbf{M}_{rc}$ , is for the transition for observing state  $r$  at the  $(v-1)^{th}$  and state  $c$  at the  $v^{th}$  visit.  $\mathbf{M}_{rc}$  itself is a  $k \times k$  square matrix for the transition from true state  $i$  to state  $j$  between 2 visits with the  $i$ th row and the  $j$ th column element as  $P_{ij} \cdot \gamma_{cj}$ .

The distribution of final disease status ( $\mathbf{Z}$ ) after first and  $m$  subsequent visits can be obtained as follows:

$$\mathbf{Z} = \mathbf{U} \times (\mathbf{W}_v)^m \quad (12)$$

The summation of the first  $k$  elements is the likelihood for the final observed state 1 and so on.

### 2.3 | Covariates incorporated into HMM

The covariates in the HMM can be considered as a function of either emitting probabilities or transition probabilities. When the emitting probability with the random component is characterized by a binary outcome, a logit link function relating the outcome to the covariates that are incorporated into the design matrix, say  $\mathbf{X}$ , of the systematic component can be written as follows:

$$\log\left(\frac{\pi}{1-\pi}\right) = \mathbf{X}\boldsymbol{\beta}. \quad (13)$$

While the emitting probabilities are derived from multinomial outcomes or the transition probabilities with  $k > 2$ , the polytomous logistic regression model can be used to model the effects of covariates with the following expression:

$$\log\left(\frac{\pi_h}{\pi_1}\right) = \mathbf{X}\boldsymbol{\beta}_h. \quad (14)$$

where  $h$  is the number of possible emitting event or number of state ( $k$ ).

For the continuous-time model, we applied the proportional hazards regression form to the instantaneous transition rates ( $q_{ij}$ ) expressed as

$$q_{ij} = q_{ij0} \cdot \exp\left(\mathbf{X}\boldsymbol{\beta}_{ij}^q\right). \quad (15)$$

where  $i, j = 1, \dots, k$ ,  $\mathbf{X}$  is the design matrix for covariates and  $\boldsymbol{\beta}_{ij}^q$  is the regression coefficient vector corresponding to  $\mathbf{X}$ .

### 3 | DESIGN AND ESTIMATION WITH BAYESIAN UNDERPINNING

#### 3.1 | 2-stage validation design on measurement errors

We proposed a Bayesian measurement-error-driven hidden Markov regression model for calibrating the effects of state-specific covariates on multistate outcomes in conjunction with a 2-stage validation design. In the first stage, participants are examined by both a gold standard and the proxy-measuring tool. Let  $n_{ab}$  denote the number of participants who are rated as states  $a$  and  $b$  by the proxy measurement and gold standard, respectively. The array of  $(n_{1b}, n_{2b}, \dots, n_{kb})$  forms the parameter of Dirichlet distribution, which generates the prior information for measurement errors.

$$\gamma_b \sim \text{Dirichlet}(n_{1b}, n_{2b}, \dots, n_{kb}), \quad b = 1, 2, \dots, k, \quad (16)$$

where  $k$  is the possible number of disease states. The special case of binomial distribution for any specific state will be applied when  $k$  is 2.

The prior information is further developed into the posterior distribution in combination with the likelihood function based on the data on a larger-scale survey at second stage.

#### 3.2 | Bayesian directed acyclic graphic model

Based on the likelihood functions derived in the previous section, the numbers of subjects classified as state 1, 2, ..., and  $k$  at the first visit, denoted by  $\mathbf{X}_F = \{O_1^{(1)}, O_2^{(1)}, \dots, O_k^{(1)}\}$ , follows a multinomial distribution with parameters of  $P(O_1^{(1)})$  (see Equation (4)),  $P(O_2^{(1)})$ , ..., and  $P(O_k^{(1)})$ , which are expressed in an array of  $\mathbf{P}_F$ . Similarly, numbers of patients in state 1, 2, ..., and  $k$  at the second visit given the state  $a$  classified in the first visit, denoted by  $\mathbf{X}_{S_a} = \{O_1^{(2)}|O_a^{(1)}, O_2^{(2)}|O_a^{(1)}, \dots, O_k^{(2)}|O_a^{(1)}\}$ , which follows multinomial distributions of parameters of  $\mathbf{P}_{S_a} = \{P(O_1^{(2)}|O_a^{(1)}), P(O_2^{(2)}|O_a^{(1)}), \dots, P(O_k^{(2)}|O_a^{(1)})\}$ .

Collectively, the full set of parameters and relevant distributions used in Bayesian MCMC simulation can be expressed as

$$\mathbf{X}_F \sim \text{Multinomial}(\mathbf{P}_F, \mathbf{N}_F), \quad (17)$$

$$\mathbf{X}_{S_a} \sim \text{Multinomial}(\mathbf{P}_{S_a}, \mathbf{N}_{S_a}). \quad (18)$$

The priors of regression coefficients are assigned as noninformative normal distribution as follows.

$$\beta' s \sim \text{Normal}(0, 10^6) \quad (19)$$

The priors of baseline transition rates are assigned as gamma distribution as follows.

$$q'_0 \sim \text{Gamma}(0.001, 0.001) \quad (20)$$

The priors of parameters on misclassifications are specified as Dirichlet distribution;

$$r_b \sim \text{Dirichlet}(c_{1b}, \dots, c_{k,b}), \quad (21)$$

where  $c_{ab}$  is the empirical number of subjects who are actually at state  $b$  based on gold standard but are classified as state  $a$  by the proxy measurement. Note that we used informative prior for misclassification terms in the light of our validated results calibrated by a gold standard.

The full joint probability distribution can then be expressed as

$$P(X_F, X_{S_a}, q'_0s, \beta' s, r' s) \propto P(X_F | q'_0s, \beta' s, r' s) \times P(X_{S_a} | q'_0s, \beta' s, r' s) \times P(q'_0s) \times P(\beta' s) \times P(r' s). \tag{22}$$

Bayesian inference using Gibbs sampling (WinBugs) program was used to derive the posterior distribution of parameters of interest when MCMC simulation is implemented. The mean values and 95% credible intervals of the proportion of misclassification, baseline transition rates, regression coefficients, and the percentage of bias can be derived to make statistical inference.

## 4 | SIMULATION FOR THE EFFECT OF MISCLASSIFICATION ON MULTISTATE OUTCOMES

### 4.1 | Differential and nondifferential misclassification

Differential and nondifferential misclassifications for these multistate outcomes of interest are defined as whether the impact of misclassifications on the effect size of covariate depends on the status of covariate. To simplify the illustration, we begin with a binary covariate (such as family history) and  $k$ -state disease model with the discrete-time hidden Markov process proposed as above.

Suppose the association between exposure and true disease status ( $T_1, T_2, \dots, T_k$ ) is displayed as follows:

	$T_1$ (State 1)	$T_2$ (State 2)	...	$T_k$ (State $k$ )
$X = 1$	$n_{11}$	$n_{12}$	...	$n_{1k}$
$X = 0$	$n_{01}$	$n_{02}$	...	$n_{0k}$

Take state 1 as the reference group, the estimated regression coefficient of  $X$  for state  $g$  in equation (14) is

$$\hat{\beta}_g = \frac{n_{1g} \cdot n_{01}}{n_{11} \cdot n_{0g}}. \tag{23}$$

In the case of misclassification, the data layout by observed and true state is presented as follows.

	$O_1, T_1$	$O_1, T_2$	...	$O_1, T_k$	...	$O_k, T_1$	$O_k, T_2$	...	$O_k, T_k$
$X = 1$	$n_{11} \cdot r_{11}   X = 1$	$n_{12} \cdot r_{12}   X = 1$	...	$n_{1k} \cdot r_{1k}   X = 1$	...	$n_{11} \cdot r_{k1}   X = 1$	$n_{12} \cdot r_{k2}   X = 1$	...	$n_{1k} \cdot r_{kk}   X = 1$
$X = 0$	$n_{11} \cdot r_{11}   X = 0$	$n_{12} \cdot r_{12}   X = 0$	...	$n_{1k} \cdot r_{1k}   X = 0$	...	$n_{11} \cdot r_{k1}   X = 0$	$n_{12} \cdot r_{k2}   X = 0$	...	$n_{1k} \cdot r_{kk}   X = 0$

The observed data layout aggregated by observed status is expressed as follows.

	$O_1 (T_1 + T_2 + \dots T_k)$	$O_2 (T_1 + T_2 + \dots T_k)$	...	$O_k (T_1 + T_2 + \dots T_k)$
$X = 1$	$\sum_{h=1}^k n_{1h} \cdot r_{1h X=1}$	$\sum_{h=1}^k n_{1h} \cdot r_{2h X=1}$	...	$\sum_{h=1}^k n_{1h} \cdot r_{kh X=1}$
$X = 0$	$\sum_{h=1}^k n_{0h} \cdot r_{1h X=0}$	$\sum_{h=1}^k n_{0h} \cdot r_{2h X=0}$	...	$\sum_{h=1}^k n_{0h} \cdot r_{kh X=0}$

The observed (uncalibrated) regression coefficient,  $\tilde{\beta}_g$ , is

$$\frac{\left( \sum_{h=1}^k n_{1h} \cdot r_{gh|X=1} \right) \cdot \left( \sum_{h=1}^k n_{0h} \cdot r_{1h|X=0} \right)}{\left( \sum_{h=1}^k n_{1h} \cdot r_{1h|X=1} \right) \cdot \left( \sum_{h=1}^k n_{0h} \cdot r_{gh|X=0} \right)}. \tag{24}$$

In the case of nondifferential misclassification, the proportion of misclassifying state  $b$  as  $a$  is independent of the exposure status,  $X$  ( $r_{ab} | X = 1 = r_{ab} | X = 0$ ). The nondifferential misclassification often lead to the problem of bias toward the null in 2-state progressive model, namely,  $|\tilde{\beta}_g| < |\hat{\beta}_g|$ .<sup>23</sup> However, it is complicated whenever the measurement errors

are involved with multistate outcomes. We therefore performed a series of simulations as shown in the following section to investigate whether such a nondifferential misclassification leads to underestimated or overestimated effect (see below).

In the case of differential misclassification, the proportion of misclassifying state  $b$  as  $a$  is dependent on the exposure status,  $X$  ( $r_{ab} | X = 1 \neq r_{ab} | X = 0$ ). The differential misclassification can result in bias in either direction in the 2-state progressive model.<sup>23</sup> However, whether differential misclassification leads to overestimation or underestimation in the misclassification of multistate outcomes is highly subject to the relative effect size of misclassification on the severity of multistate disease outcome. This will be demonstrated by the following simulated results.

The percentage of the bias due to measurement errors for state  $g$  is defined as  $P_{bias} = \frac{\tilde{\beta}_g - \hat{\beta}_g}{|\hat{\beta}_g|} \times 100\%$ . For the case of risk factors, a positive percentage refers to overestimation, and a negative percentage refers to underestimation.

## 4.2 | Simulation for differential types of misclassification

In this section, we simulated various degrees of misclassifications of disease status in various exposure groups (such as with or without family history) to investigate how differential and nondifferential misclassification led to underestimation or overestimation of effect size.

We developed a simulated algorithm to elucidate the effect of misclassification on the effect size of the covariate of interest. For ease of illustration, we begin with a binary risk factor (exposed versus unexposed), say  $X$ , and 2-state outcome (disease [state 2] versus nondisease [state 1]). The proportion of having the exposure ( $X = 1$ ) of covariate was 30%. The prevalence of disease in the exposed and unexposed group was 30% and 20%, respectively, which yielded a true estimated odds ratio of 1.7143 ( $\hat{\beta} = 0.5390$ ).

We used the Monte Carlo Markov chain method to obtain the distribution of measurement error. The simulation procedure for a 2-state model is described below.

1. For the proportion of misclassification between state 1 and state 2 ( $r_{12}$  and  $r_{21}$ ), we independently drew random numbers from uniform distribution (0, 0.5) for the exposed ( $r_{12} | X = 1$  and  $r_{21} | X = 1$ ) and unexposed groups ( $r_{12} | X = 0$  and  $r_{21} | X = 0$ ) following the notations indicated above.
2. Following the decomposition methods in the methods section, we estimated the associated observed regression coefficient and the corresponding percentage of bias.

It should be noted that the procedure mentioned above was for differential misclassification. For nondifferential misclassification that is independent of the status of covariates, we simply drew random numbers of  $r_{21} | X = 1$  and  $r_{12} | X = 1$  as described in step 1, and assigned  $r_{21} | X = 0 = r_{21} | X = 1$  and  $r_{12} | X = 0 = r_{12} | X = 1$ .

For the 3-state disease, we used a hypothetical cohort in which the prevalence of state 2 and state 3 disease was 20% and 10% in the exposed group, respectively, and 15% and 5% in the unexposed group, respectively, for illustration. The odds ratio ( $\Omega$ ) of exposure for being state 2 and state 3 versus state 1 was  $\Omega_{21} = 1.5238$  ( $\hat{\beta}_{21} = 0.4212$ ), and  $\Omega_{32} = 2.2857$  ( $\hat{\beta}_{32} = 0.8267$ ), respectively.

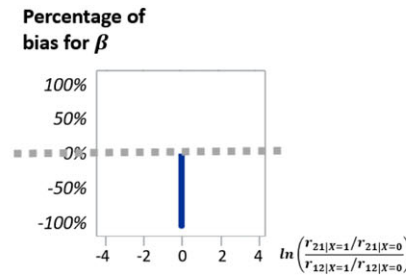
The simulated results regarding the relationship between  $\ln\left(\frac{r_{ab|X=1}/r_{ab|X=0}}{r_{ba|X=1}/r_{ba|X=0}}\right)$  ( $a, b \in 1, 2, \text{ and } 3$ ) and the percentage of bias as indicated above are presented for the nondifferential misclassification and the differential misclassification of 2-state and 3-state model. Based on the simulated results, the lower and upper bounds of  $\ln\left(\frac{r_{ab|X=1}/r_{ab|X=0}}{r_{ba|X=1}/r_{ba|X=0}}\right)$  corresponding to 1% to 99% underestimation could be derived.

## 4.3 | Simulated results

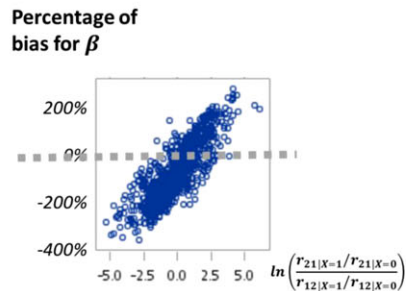
We first present the simulated results for a 2-state disease process. Given the nondifferential misclassification with 10 000 repeats, all samples led to underestimation ( $P_{bias} < 0$ ) (Figure 1A). When the misclassification was differential, 33.11% of



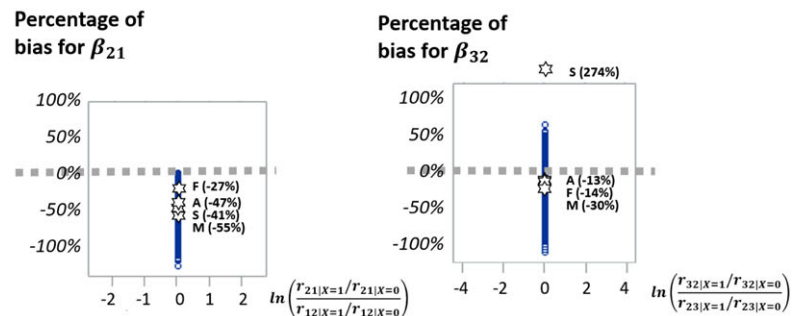
(A)



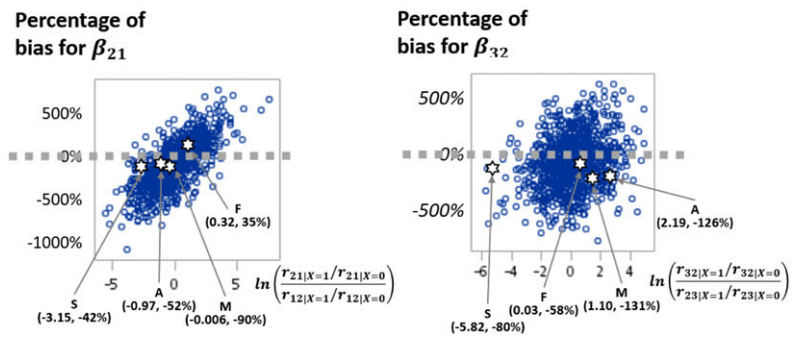
(B)



(C)



(D)



**FIGURE 1** Simulated results of the percentage of biases in 2-state and 3-state disease model involved with nondifferential and differential misclassification. A, 2-state model with nondifferential misclassification; B, 2-state model with differential misclassification; C, 3-state model with nondifferential misclassification; and D, 3-state model with differential misclassification. \*The stars in Figure 1C,D indicated the plots of  $\ln\left(\frac{r_{ab|X=1}/r_{ab|X=0}}{r_{ba|X=1}/r_{ba|X=0}}\right)$  to percentage of bias for the androgenetic alopecia example with associated indicating letters for different covariates: M for metabolic syndrome, A for age group, S for sex, and F for family history [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

samples were overestimation. Interestingly, the proportion of overestimation increased when  $\ln\left(\frac{r_{21|X=1}/r_{21|X=0}}{r_{12|X=1}/r_{12|X=0}}\right)$  increased (Figure 1B). The proportion of overestimation was 4.78% when  $\ln\left(\frac{r_{21|X=1}/r_{21|X=0}}{r_{12|X=1}/r_{12|X=0}}\right)$  was negative, whereas the corresponding figures increased to 23.78% when  $\ln\left(\frac{r_{21|X=1}/r_{21|X=0}}{r_{12|X=1}/r_{12|X=0}}\right)$  was between 0 and  $\beta$  (0.5390) and 76.81% when

greater than  $\beta$ , respectively. Samples with this index smaller than  $-0.69$  or larger than  $1.95$  were associated with 1% and 99% overestimation, respectively.

In the scenario of the nondifferential misclassification for a 3-state disease, underestimation was noted for 98.8% and 87.6% of random samples of  $\beta_{21}$  and  $\beta_{32}$ , respectively (Figure 1C). Given the differential misclassification, 38.76% of samples were overestimation for  $\beta_{21}$ , which increased with the elevated value of  $\ln\left(\frac{r_{21|X=1}/r_{21|X=0}}{r_{12|X=1}/r_{12|X=0}}\right)$  (Figure 1D). When

$\ln\left(\frac{r_{21|X=1}/r_{21|X=0}}{r_{12|X=1}/r_{12|X=0}}\right)$  was negative, the proportion of overestimation was 13.06%. The corresponding proportions for  $\ln\left(\frac{r_{21|X=1}/r_{21|X=0}}{r_{12|X=1}/r_{12|X=0}}\right)$  between 0 and  $\beta_{21}(=0.4212)$  and greater than  $\beta_{21}$  were 35.68% and 72.24%, respectively. Samples with  $\ln\left(\frac{r_{21|X=1}/r_{21|X=0}}{r_{12|X=1}/r_{12|X=0}}\right)$  smaller than  $-1.32$  or larger than  $2.17$  yielded 1% and 99% overestimation, respectively. As far as  $\beta_{32}$  is concerned, there were 33.20% sample resulting in overestimation, which also increased when  $\ln\left(\frac{r_{32|X=1}/r_{32|X=0}}{r_{23|X=1}/r_{23|X=0}}\right)$  increased (Figure 1D). The corresponding proportions of overestimation when  $\ln\left(\frac{r_{32|X=1}/r_{32|X=0}}{r_{23|X=1}/r_{23|X=0}}\right)$  less than 0, between zero and  $\beta_{32}(=0.8267)$ , and greater than  $\beta_{32}$  were 24.54%, 32.37%, and 44.26%, respectively. Samples with  $\ln\left(\frac{r_{32|X=1}/r_{32|X=0}}{r_{23|X=1}/r_{23|X=0}}\right)$  smaller than  $-2.67$  or larger than  $3.32$  yielded 1% and 99% overestimation, respectively.

In summary, a series of simulation algorithms were developed to assess underestimation and overestimation of effect size due to the misclassification of multistate outcome. The 95% credible interval for biased direction, particularly overestimation, can be obtained given various scenarios.

## 5 | APPLICATION TO AGA

Androgenetic alopecia is a common health problem. It is a progressive disease with a cascade of multiple-step progressions by using different classifications. For example, Norwood system classifies AGA into 7 categories from mild to extremely severe types.<sup>24</sup> In the light of this inherent progressive property, it is of great interest to quantify the instantaneous rate of progression from free of AGA until severe AGA in the province of dermatology by using a stochastic process. However, such a natural history of AGA progression has been barely addressed.

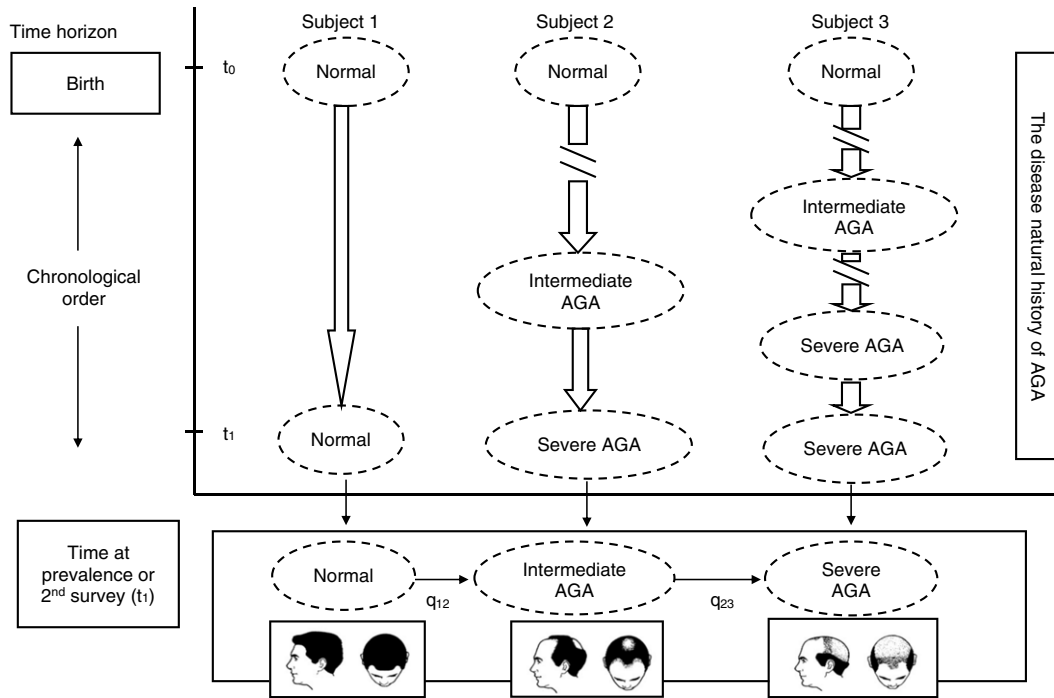
In addition to the disease natural history of AGA, it is also interesting to assess risk factors responsible for different stages of AGA. The AGA has been recently reported to be associated with metabolic syndrome (MetS).<sup>25</sup> However, whether the contribution of MetS to AGA plays a crucial role in the outset of AGA or a promoter for progression to severe AGA is elusive and worthy of being investigated.

The classification of AGA is prone to misclassification errors when multistate outcomes are measured by nondermatologist such as public health nurses in a large community-based survey. Owing to inherent progressive property, the role of MetS in association with AGA progression, and classification errors, we applied the proposed measurement-error-driven HMM with Bayesian underpinning to estimate transition rates between AGA states. The respective effects of age, sex, family history and MetS on each transition step were also modeled by considering classification error rates.

Although there are 2 types of misclassification as mentioned in the simulated section, we feel that nondifferential misclassification is more likely to be encountered in the current example of AGA than differential misclassification because the public health nurses had no information on MetS of examinees when they rated the degree of AGA.

### 5.1 | Model specification of 3-state AGA

We used a 3-state continuous-time Markov model for the progression of grades in AGA with a state space  $\Omega=\{1, 2, 3\}$ , where 1 represents normal (normal or Norwood type I), 2 for intermediate AGA (Norwood type II-IV and Ludwig L-I), and 3 for severe AGA (Norwood type V-VII and Ludwig L-II and L-III). Figure 2 shows the disease natural history model for depicting the progression of AGA. In this illustration, all 3 subjects start from normal in hair status at birth or the first time of survey. Subject 1 stays in the normal state until the time of the first survey or subsequent surveys. Subject 2 progresses to intermediate AGA at the time between birth and prevalence survey or between the first and



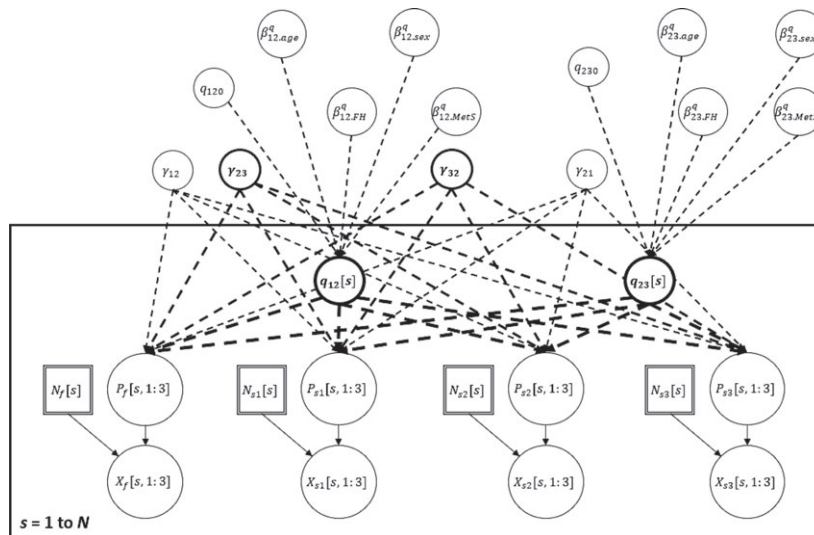
**FIGURE 2** The disease natural history model in androgenetic alopecia (AGA)

second survey. Likewise, subject 3 progresses to intermediate and severe AGA before the first survey. Each subject of the underlying population would follow 1 of 3 pathways. In Figure 2, annual progression rates in the 3-state model are denoted by  $q_{12}$  and  $q_{23}$ , representing 2 progressions from normal to intermediate and from intermediate to severe AGA. To incorporate effects of covariates, such as age, sex, family history of AGA, and MetS, we analyzed data with the proportional hazards regression form following the treatment of continuous-time Markov model in the methods section (Equation (15)). The Bayesian directed acyclic graphic (DAG) model for the 3-state model considering misclassifications is illustrated in Figure 3. By using Equation (15), the transition rates for the  $s$ th covariate profile,  $q_{12}[s]$  and  $q_{23}[s]$ , were modeled as functions of covariates (age[s], sex[s], FH[s], and MetS[s]) and their corresponding regression coefficients,  $(\beta_{12.age}^q, \beta_{12.sex}^q, \beta_{12.FH}^q, \beta_{12.MetS}^q, \beta_{23.age}^q, \beta_{23.sex}^q, \beta_{23.FH}^q, \text{ and } \beta_{23.MetS}^q)$ . By using Equations (2) to (15), probabilities of observing cases of the  $s$ th covariate profile in different states in the first ( $P_f[s, 1:3]$ ) and the second visit ( $P_{s1}[s, 1:3]$ ,  $P_{s2}[s, 1:3]$  and  $P_{s3}[s, 1:3]$ ) were the function of measurement error ( $\gamma_{12}, \gamma_{23}, \gamma_{21}, \gamma_{32}$ ) and transition rates ( $q_{12}[s]$  and  $q_{23}[s]$ ). Note that the links for the logic function are illustrated with dashed arrow in the DAG model. As far as the stochastic link is concerned, the random variables  $X_f[s, 1:3]$ , an array with numbers at states 1 to 3 in the first visit, follow a multinomial distribution with parameters of probabilities ( $P_f[s, 1:3]$ ) and total number of subjects of the  $s$ th covariate profile ( $N_f[s]$ ) by using Equation (17) and the link is illustrated with solid arrow. Similarly, the random variables  $X_{s1}[s, 1:3]$ - $X_{s3}[s, 1:3]$  are linked with their associated parameters according to Equation (18).

### 5.2 | Study population involved in the 2-stage validation design

In this study, we conducted a community-based 2-stage validation study design for the effect of covariates on AGA considering measurement errors. In the first stage, a calibration study was conducted with AGA categorized by both a senior dermatologist (gold standard) and public health nurses on 555 subjects to provide information on measurement errors because those public health nurses would be responsible for rating the severity of AGA in the main study at second stage.

In the main study, the study population was derived from a community-based integrated screening program in Tainan County, Taiwan. Two surveys for AGA were conducted between 2005 and 2010. A total of 7960 subjects aged 40 years or older who attended the screening program in 2005 were invited to have the first community-based AGA survey. A total of 6817 subjects had complete data on the result of AGA survey and all other screening items. The second survey was conducted in 2010 for 49 936 subjects attending the screening program. Among them, 42 000 subjects had



Annotation:

- (1)  $N$ : number of covariate profiles.
- (2)  $N_j[m]$ ,  $N_{s1}[m]$ ,  $N_{s2}[m]$  and  $N_{s3}[m]$ : number of subjects for the  $m^{\text{th}}$  covariate profile in the first visit, and in states 1-3 from the previous visit.
- (3)  $X_j[m, 1:3]$ ,  $X_{s1}[m, 1:3]$ ,  $X_{s2}[m, 1:3]$  and  $X_{s3}[m, 1:3]$ : number of subjects in state 1-3 for the  $m^{\text{th}}$  covariate profile in the first visit, and of those in states 1-3 from the previous visit, which follows the multinomial distribution.
- (4)  $q_{ij}[s]$ : the transition rate from state  $i$  to state  $j$  for the  $s^{\text{th}}$  covariate profile.
- (5)  $q_{ij0}$ : the transition rate from state  $i$  to state  $j$  for the covariates at baseline value.
- (6)  $\gamma_{ij}$ : the conditional probability for those with true state  $j$  and classified as state  $i$ .
- (7)  $\beta_{ij,X}^q$ : the regression coefficient of covariate  $X$  for the transition rate  $q_{ij}$  from state  $i$  to state  $j$ .

**FIGURE 3** The acyclic graph model for the multistate model incorporating measurement errors

complete data on the survey of AGA and all screening items. There were 1440 subjects participating in both surveys who are available for estimating the incidence of intermediate AGA and the progression rate from intermediate to severe AGA and to investigate the effects of covariates of interest on these progression rates.

### 5.3 | Data collection

Androgenetic alopecia state was classified by public health nurses who had taken the training course by a senior dermatologist before the survey in the first stage of the study. On the screening site, public health nurses also took the anthropometric measures (for body height, body weight, and circumferences of waist and hip) and blood pressure readings. Subjects were requested to fast for at least 8 hours before screening for the biochemical examination, including fasting glucose, lipid profile, liver function, etc. Data on lifestyle, personal disease history, and family history were collected with a structured questionnaire. The written informed consent has been obtained from each participant.

Androgenetic alopecia involving frontal hairline was classified according to the Norwood classification, grading from type I to VII. Ludwig classification, grading from L-I to L-III, was applied if frontal hairline was not involved.<sup>26</sup>

We used NCEP ATP III criteria to define MetS.<sup>27</sup> Briefly, subjects met at least 3 of the following criteria were classified as having MetS: (1) central obesity (waist circumference larger than or equal to 80 cm for female and 90 cm for male, in the light of Asian modifications),<sup>28</sup> (2) hypertriglyceride (triglyceride higher than or equal to 150 mg/dL), (3) hypohigh density lipoprotein cholesterol (HDL-C) (HDL-C less than 50 mg/dL for female and less than 40 mg/dL for male), (4) elevated blood pressure (systolic blood pressure is 130 mmHg or above or diastolic blood pressure is 85 mmHg or above), and (5) hyperglycemia (fasting glucose is 100 mg/dL or above).

### 5.4 | Estimated results

Table 1 shows the results of calibrating AGA rated by public health nurses compared with the senior dermatologist. Among the 185 subjects whom the senior dermatologist classified as AGA free, 7.6% ( $n = 14$ ) were misclassified as intermediate AGA. Among the 222 subjects with intermediate AGA categorized by gold standard, the erroneous rates of downgrading as AGA free and upgrading as severe AGA were 2.3% ( $n = 5$ ) and 3.6% ( $n = 8$ ), respectively. Interestingly, the misclassification rate was highest in the severe AGA; there were 16.2% ( $n = 24$ ) misclassification among the 148 severe subjects.

**TABLE 1** Comparison of androgenetic alopecia classification rated by public health nurses and the senior dermatologist (gold standard)

Gold Standard	Public Health Nurses			Total
	Normal	Intermediate	Severe	
Normal	171 (92.4%)	14 (7.6%)	0	185
Intermediate	5 (2.3%)	209 (94.1%)	8 (3.6%)	222
Severe	0	24 (16.2%)	124 (83.8%)	148

Table S1 shows the estimated results of univariate analysis on the effect of each covariate on the occurrence of intermediate AGA and also on subsequent progression to severe AGA with and without considering measurement errors. It can be clearly seen that age, sex, and family history made contributions not only to the occurrence of intermediate AGA but also to subsequent progression. It should be noted that considering measurement errors as opposed to the neglect of measurement errors rendered all 3 estimates inflated, which implies the underestimation of the impacts of all these 3 factors on 2 step-by-step progressions without making allowance for the measurement errors. The effects of MetS on 2 transitions were not statistically significant. As far as individual components are concerned, elevated fasting sugar and hypertension were 2 statistically significant factors responsible for 2 transitions. Like age, sex, and family history, the effects of these 2 components were also underestimated without considering measurement errors. Three other components had various contributions to 2 transitions.

Table 2 shows the estimated regression coefficient of MetS on the occurrence and progression of AGA after adjusting for age, sex, and family history of AGA before considering measurement errors. The results show that MetS plays a more significant role in the progression (transition from intermediate to severe state) than the occurrence of AGA (transition from normal to intermediate state). The magnitude of the effect (relative risk [RR], the exponential transform of

**TABLE 2** Estimated baseline transition rates and regression coefficients of metabolic syndrome (MetS) and its individual components after adjusting for age, sex, and family history

	Normal to Intermediate			Intermediate to Severe		
	Estimate	SD	95% CI	Estimate	SD	95% CI
Baseline transition rates	0.0021	0.00004	0.0021-0.0022	0.0030	0.0002	0.0026- 0.0033
MetS						
Yes vs no	0.0175	0.0268	-0.0350 to 0.0710	0.1478	0.0585	0.0295-0.2631
Age						
>70 vs ≤70	0.1645	0.0219	0.1233-0.2071	0.2310	0.0468	0.1384-0.3255
Sex						
Male vs female	1.2660	0.0201	1.2270-1.2660	0.8491	0.0591	0.7328-0.9666
Family history						
Yes vs no	1.0550	0.0480	0.9586, 1.1480	0.6911	0.0849	0.5275-0.8569
Models for individual MetS components						
High waist circumference						
>80 cm (female), >90 cm (male)	-0.0162	0.0207	-0.0563 to 0.0250	0.0140	0.0493	-0.0825 to 0.1103
Elevated triglyceride						
≥150 mg/dL	0.0060	0.0228	-0.0395 to 0.0501	0.0425	0.0530	-0.0617 to 0.1458
Low level of high density lipoprotein						
<50 mg/dL (female), <40 mg/dL (male)	0.0440	0.0280	-0.0113 to 0.0995	0.0492	0.0685	-0.0854 to 0.1813
Elevated fasting glucose						
≥110 mg/dL or Diabetes Mellitus (DM)	0.0626	0.0237	0.0162-0.1088	0.1812	0.0534	0.0730-0.2836
Hypertension						
Yes vs no	0.0429	0.0199	0.0043-0.0823	0.1519	0.0456	0.0636-0.2409

regression coefficients) of MetS was 1.16 ( $=\exp(0.1478)$ ) (95% CI, 1.03-1.30) for the progression and 1.02 (95% CI, 0.97-1.07) for the occurrence of AGA. Table 2 also shows the effect of the individual components of MetS. Elevated fasting glucose and hypertension relative to 3 other components made significant contributions to occurrence of AGA and subsequent progression from intermediate to severe state. The RRs of elevated fasting glucose on the occurrence and progression of AGA were 1.06 (95% CI, 1.02-1.11) and 1.20 (95% CI, 1.08-1.33), respectively. The corresponding figures for hypertension were 1.04 (95% CI, 1.00-1.09) and 1.16 (95% CI, 1.07-1.27).

Table 3 shows the results with Bayesian approach incorporating informative prior for the measurement errors borrowing from Table 1 given the nondifferential misclassification that may be adequate for this AGA example as indicated above. The misclassifications were noted more prominent in misclassifying state 2 to state 1 ( $\widehat{\gamma}_{12} = 19.5\%$ , 95% CI, 15.6%-23.6%) and state 3 to state 2 ( $\widehat{\gamma}_{23} = 14.8\%$ , 95% CI, 10.1%-20.1%). Misclassifications from lower to higher states seemed to be less prominent. They were 8.1% (95% CI, 7.2%-8.9%) and 8.6% (95% CI, 6.3%-10.9%) for misclassifying state 1 to 2 ( $\widehat{\gamma}_{21}$ ) and state 2 to 3 ( $\widehat{\gamma}_{32}$ ), respectively. Metabolic syndrome was still statistically significantly associated with the progression of AGA in the second transitions (RR = 1.23, 95% CI, 1.00-1.50), but a lacking of significant association with the occurrence of AGA was found (RR = 1.04, 95% CI, 0.96-1.13). Similarly, the individual components of elevated fasting glucose and hypertension still play significant roles in both transitions as shown in the model without

**TABLE 3** Estimated measurement errors, baseline transition rates, and regression coefficients of metabolic syndrome (MetS) and its individual components after adjusting for age, sex, and family history by Bayesian approach with informative prior

	Normal to Intermediate			Intermediate to Severe		
	Estimate	SD	95% CI	Estimate	SD	95% CI
Proportion of misclassifications						
$\gamma_{12}$	19.5%	0.0205	15.6%-23.6%			
$\gamma_{21}$	8.1%	0.0042	7.2%-8.9%			
$\gamma_{23}$	14.8%	0.0254	10.1%-20.1%			
$\gamma_{32}$	8.6%	0.0117	6.3%-10.9%			
Baseline transition rates	0.0009	0.0001	0.0007-0.0011	0.0043	0.0011	0.0024-0.0068
MetS						
Yes vs no	0.0407	0.0410	-0.0418 to 0.1206	0.2078	0.1034	0.0042-0.4071
Age						
>70 vs $\leq 70$	0.3126	0.0331	0.2478-0.3787	0.2676	0.0920	0.0957-0.4609
Sex						
Male vs female	2.1320	0.0877	1.9740-2.3130	0.1332	0.1969	-0.2414 to 0.5158
Family history						
Yes vs no	1.4420	0.0771	1.2900-1.5950	0.8033	0.1404	0.5373-1.0770
Models for individual MetS components						
High waist circumference						
>80 cm (female), >90 cm (male)	0.0021	0.0318	-0.0612 to 0.0644	0.0024	0.0880	-0.1698 to 0.1755
Elevated triglyceride						
$\geq 150$ mg/dL	0.0220	0.0337	-0.0439 to 0.0864	0.0328	0.0913	-0.1501 to 0.2107
Low level of high density lipoprotein						
<50 mg/dL (female), <40 mg/dL (male)	0.0403	0.0460	-0.0492 to 0.1282	0.0783	0.1184	-0.1567 to 0.3028
Elevated fasting glucose						
$\geq 110$ mg/dL or DM	0.0955	0.0368	0.0232-0.1670	0.2189	0.0881	0.0435-0.3918
Hypertension						
Yes vs no	0.0703	0.0302	0.0110-0.1300	0.2348	0.0824	0.0772-0.3998

considering measurement errors (Table 2). However, the magnitude of the effect was away from null hypothesis further when measurement errors were taken into account. The RRs of fasting glucose on the occurrence and progression of AGA became 1.10 (95% CI, 1.02-1.18) and 1.24 (95% CI, 1.04-1.48), respectively. The corresponding figures for hypertension were 1.07 (95% CI, 1.01-1.14) and 1.26 (95% CI, 1.08-1.49), respectively. It should be noted that the standard deviation of the posterior distribution was larger in the measurement error model compared with the uncalibrated ones (Table 2).

The underestimation resulting from the neglect of measurement errors were observed for all covariates on the occurrence of AGA as shown in Table 4 and Figure 1C with negative biased percentage for all the variables except HDL. The percentage of bias was statistically significant for age, sex, and family history but not for MetS or its individual components (Table 4). As shown in the right panel of Figure 1C, the underestimation of the regression coefficient ( $\beta_{23}$ ) on the progression of AGA was not so robust as that of the regression coefficient ( $\beta_{12}$ ) on the incidence of AGA (left panel of Figure 1C). The percentage of bias was not significant for all the covariates on the progression of AGA, except an overestimation for sex. It should be also noted that the measurement error model with informative prior had better model performance than the one without considering no misclassification (Table 2). The estimate of Deviance Information Criterion (DIC) for the latter was 6428, but it was reduced to 3774 for the former calibrated model.

We also tested the measurement error model with noninformative prior (Table 5). The estimates of misclassifying toward less severe stage were exaggerated ( $\hat{\gamma}_{12} = 55.6\%$ , 95% CI, 35.7%-65.7%;  $\hat{\gamma}_{23} = 50.9\%$ , 95% CI, 25.9%-59.8%) compared with their counterparts in the model with informative prior (Table 3), whereas the estimates of misclassifying toward severe state were decreased ( $\hat{\gamma}_{21} = 4.7\%$ , 95% CI, 2.2%-7.3%;  $\hat{\gamma}_{32} = 4.0\%$ , 95% CI, 2.5%-6.9%). The standard deviations of the posterior distribution were further enlarged compared with the corresponding figures of the model with informative prior for measurement error. Together with the fact that the point estimates of the regression coefficients of MetS on both transitions were shrunk, MetS became an insignificant factor responsible for both occurrence (RR = 1.02, 95% CI, 0.91-1.14) and the progression of AGA (RR = 1.16, 95% CI, 0.90-1.48) (Table 5). We also found that

**TABLE 4** Percentage of biases of regression coefficients of metabolic syndrome (MetS) and its individual components due to misclassification

	Normal to Intermediate		Intermediate to Severe	
	Median	95% CI	Median	95% CI
MetS				
Yes vs no	-55	-262 to 953	-30	-87 to 553
Age				
>70 vs ≤70	-47	-63 to -27	-13	-60 to 160
Sex				
Male vs female	-41	-46 to -35	274	87-845
Family history				
Yes vs no	-27	-36 to -16	-14	-42 to 36
Models for individual MetS components				
High waist circumference				
>80 cm (female), >90 cm (male)	-108	-1796 to 301	40	-457 to 930
Elevated triglyceride				
≥150 mg/dL	-61	-540 to 954	21	-249 to 1595
Low level of high density lipoprotein				
<50 mg/dL (female), <40 mg/dL (male)	91	-126 to 3494	30	-337 to 1918
Elevated fasting glucose				
≥110 mg/dL or DM	-35	-84 to 201	-17	-71 to 320
Hypertension				
Yes vs no	-39	-94 to 301	-35	-77 to 113

**TABLE 5** Estimated measurement errors, baseline transition rates, and regression coefficients of metabolic syndrome (MetS) and its individual components after adjusting for age, sex, and family history by Bayesian approach with noninformative prior

	Normal to Intermediate			Intermediate to Severe		
	Estimate	SD	95% CI	Estimate	SD	95% CI
Proportion of misclassifications						
$\gamma_{12}$	55.6%	0.0749	35.7%-65.7%			
$\gamma_{21}$	4.7%	0.0130	2.2%-7.3%			
$\gamma_{23}$	50.9%	0.0816	25.9%-59.8%			
$\gamma_{32}$	4.0%	0.0112	2.5%-6.9%			
Baseline transition rates	0.0034	0.0011	0.0013-0.0055	0.0014	0.0008	0.0004-0.0035
MetS						
Yes vs no	0.0187	0.0581	-0.0934 to 0.1339	0.1483	0.1261	-0.1033 to 0.3931
Age						
>70 vs ≤70	0.2003	0.0708	0.069-0.3421	0.4033	0.1106	0.1815-0.6222
Sex						
Male vs female	1.6450	0.1783	1.3450-2.0380	1.5130	0.5316	0.3881-2.5260
Family history						
Yes vs no	0.8486	0.2277	0.4789-1.3650	1.8020	0.3315	1.0220-2.3860
Models for individual MetS components						
High waist circumference						
>80 cm (female), >90 cm (male)	-0.0057	0.0444	-0.092 to 0.0825	-0.0014	0.0980	-0.1936 to 0.1878
Elevated triglyceride						
≥150 mg/dL	0.0045	0.0478	-0.0934 to 0.0996	0.0588	0.1026	-0.1382 to 0.2567
Low level of high density lipoprotein						
<50 mg/dL (female), <40 mg/dL (male)	0.1291	0.0607	0.0123-0.2527	-0.1544	0.1332	-0.4228 to 0.1057
Elevated fasting glucose						
≥110 mg/dL or DM	0.1118	0.0598	-0.0003 to 0.2347	0.0893	0.1230	-0.1509 to 0.3244
Hypertension						
Yes vs no	0.0478	0.0409	-0.0314 to 0.1260	0.2017	0.0970	0.0127-0.3910

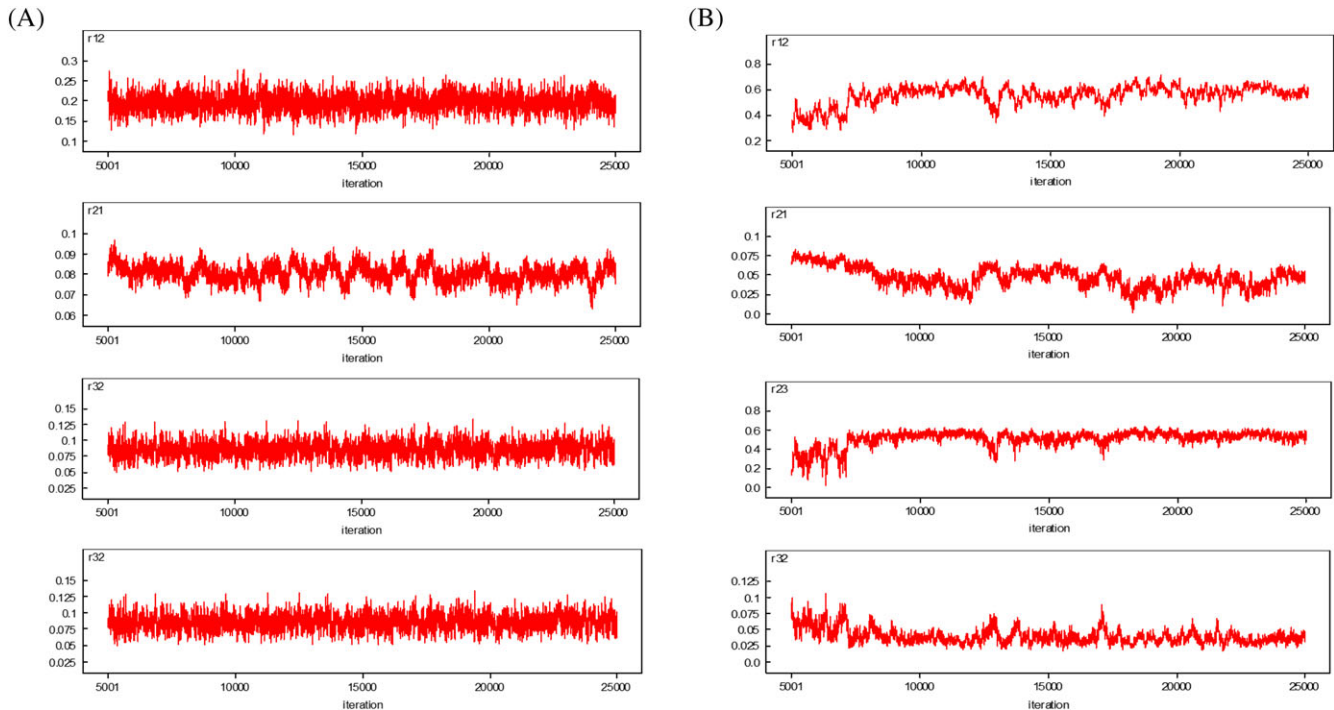
the effects of individual components on the transitions of AGA may be changed toward or away from null hypothesis or even be changed in terms of the direction of effect. For example, the effect of abnormal HDL on the progression of AGA deteriorated using the model with informative prior but ameliorated using the model with noninformative prior.

Without prior information, whether the changes of disease states in different time were due to disease progression of natural course or misclassification was indistinguishable. The model would encounter the problem of identifiability. In the current study, the model with noninformative prior resulted in negative estimate of the effective number of parameters (pD). The convergence of measurement error terms was much better improved in the model with informative prior compared with that with noninformative prior (Figure 4).

## 6 | DISCUSSION

We developed a Bayesian measurement-error-driven HMM for assessing the direction of misclassification implicated in multistate outcomes that may affect the effect of covariates on the transitions between multiple states. Information on the priors of measurement errors with Bayesian underpinning was obtained from a 2-stage validation design. There are several unique and novel statistical thoughts on the methodological development and the application to an example of AGA.





**FIGURE 4** The iterative history of sampling for measurement errors with and without informative prior in the multivariable model for metabolic syndrome [Colour figure can be viewed at wileyonlinelibrary.com]

First, the use of HMM is very flexible for modeling the misclassifications of multistate outcomes in epidemiology in contrast to one of previous specific statistical methods based on the matrix approach under the context of positive predictive value, which are restricted to the 2-state outcome and only amendable to irreversible outcome.<sup>29</sup> The proposed recursive relationships between repeated equal or unequal spaced visits with HMM may accommodate different and multiple types of measurement errors in relation to the evolution of multistate disease process. The proposed HMM can also accommodate time-dependent misclassification provided information is sufficient to be observed.

The second advantage of the proposed model that is different from the other HMM approach dealing with the misclassification of outcome<sup>30</sup> is the further assessment of direction of biased effect size of each covariate attributed to differential or nondifferential misclassification of multistate outcomes. The simulated algorithm provides a new insight into underestimation or overestimation of effect size when different scenarios of misclassification occur. Our simulated results show the majority of estimated odd ratios (ORs) are underestimated in 2-state or 3-state Markov model given a nondifferential misclassification. When the differential misclassification is implicated, whether state-specific ORs were underestimated or overestimated is highly dependent on relative size of measurement error between the exposed group and the nonexposed group. It is interesting to find that as far as a risk factor is concerned when the misclassification toward a severe state tends to occur more frequently in the exposed group than the unexposed group ( $\ln\left(\frac{r_{ab|X=1}/r_{ab|X=0}}{r_{ba|X=1}/r_{ba|X=0}}\right) > 0$ ), the estimated effect size would be more likely to be pulled away from the null (overestimation), which is more commonly seen in previous epidemiological studies. It should be noted that the patterns of overestimation and underestimation are very complicated in the measurement errors implicated in the multistate outcomes by whether they are subject to downstaging or upstaging measurement error.

Although we reckon differential misclassification may be not possible in the AGA example, we are here to present the illustration on how the proposed model can be applied to our AGA model. The estimated results of the index for misclassification as pointed out in Figure 1D show the estimates of  $\ln\left(\frac{r_{21|X=1}/r_{21|X=0}}{r_{12|X=1}/r_{12|X=0}}\right)$  for MetS, age group, sex, and family history were  $-0.0061$ ,  $-0.9707$ ,  $-3.1450$ , and  $0.3200$  with the corresponding percentage of bias being  $-89.7\%$ ,  $-51.6\%$ ,  $-42.3\%$ , and  $35.2\%$ , respectively, which indicated that the uncalibrated effect size for the transition from normal to intermediate AGA would more likely be underestimated, namely, negative percentage of bias for  $\beta_{21}$ , except for family history. Similar for  $\beta_{32}$ , the estimates of  $\ln\left(\frac{r_{32|X=1}/r_{32|X=0}}{r_{23|X=1}/r_{23|X=0}}\right)$  for MetS, age group, sex, and family history were

1.1023, 2.1948,  $-5.8150$ , and  $0.0266$  with the corresponding negative percentage of bias being  $-131.3\%$ ,  $-126\%$ ,  $-79.6\%$ , and  $-57.8\%$ , respectively, which indicated that the uncalibrated effect size, except for sex, would have half chance to be underestimated. However, the estimation based on differential misclassification was not so stable possibly because our prior information in the pilot study was not covariate dependent so that it was not included in the main results. Recall that the nondifferential misclassification is therefore the probable scenario of our AGA example as the public health nurses were not aware of the status of covariates when they rated the AGA status in the community-based survey.

Thirdly, we have applied this 3-state measurement-error-driven HMM with Bayesian DAG method to assess the effect of covariates on multistep natural progression of AGA considering the probable misclassification measured in the community. After making allowance for the errors of misclassification, it is obvious that the effect of MetS on the occurrence of intermediate stage of AGA and on the transition from intermediate to severe state has been inflated from 1.02 to 1.04 for the first transition and from 1.16 to 1.23 for the second transition corresponding to the uncalibrated model and the calibrated model one. The breakthrough of this study is to evaluate the impact of joint influence of such a misclassification on the multivariate outcome instead of only the univariate outcome. The estimates of multistep progressive of AGA and the effects of covariates enabled us to elucidate whether the associated risk factors play the role at onset of AGA or as a promoter for progression to severe AGA. Earlier, Su and Chen<sup>25</sup> found that MetS was associated with a 67% elevated risk of the presence of AGA with a cross-sectional community-based survey. In our study, we further demonstrate that MetS plays a more important role in a promoter for AGA progression (RR = 1.23, 95% CI, 1.00-1.50) than the onset of AGA (RR = 1.04, 95% CI, 0.96-1.13). Among the individual components of MetS, elevated fasting glucose and blood pressure were both statistically significant of being initiator and promoter. Such results after the calibration of measurement errors make great contribution to the identification of significant covariates played in personalized medicine of disease progression with multistate outcomes.

The fourth contribution to this theme is the better use of Bayesian approach particularly when a 2-stage validation design is applied. It can be found from our application of AGA example that lacking of prior information on measurement errors without using a 2-stage validation design led to unstable and unreliable estimation of parameters. In our example of AGA, the prior knowledge on probable misclassification was obtained in an earlier but smaller scale of calibration with public health nurses who practiced the field survey and a senior dermatologist (who were treated as a gold standard). The prior information was further incorporated by using the Bayesian approach. The results showed the estimates of measurement error may be affected and the convergence of estimation has become much stable without the manifestation of autocorrelation when using informative prior.

## 7 | COMPARISONS WITH ARTIFICIAL CONSTRAINT AND RELABELING ALGORITHM

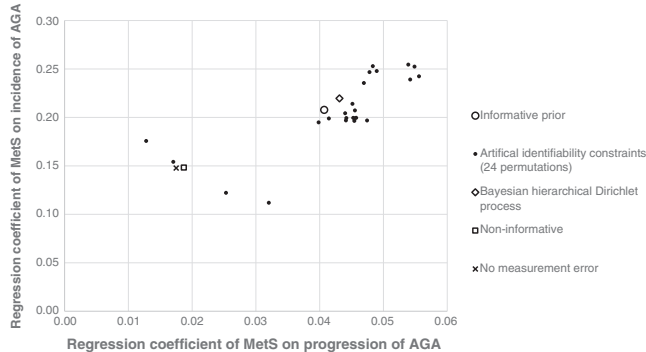
The comparisons were made between the results based on our Bayesian 2-stage method and those of both the artificial identifiability constraints approach and relabeling algorithms used in the mixed HMM without using a 2-stage design. We estimated parameters with all possible permutations of the 4 measurement error terms ( $\gamma_{12}$ ,  $\gamma_{21}$ ,  $\gamma_{23}$ , and  $\gamma_{32}$ ) based on the main dataset (the detailed estimated results shown online at website [http://my2.tmu.edu.tw/blog.php?user=amyen&f=blog\\_doc&bid=136030](http://my2.tmu.edu.tw/blog.php?user=amyen&f=blog_doc&bid=136030)). The smallest DIC (3647.22) was seen in the model with the following constraint:  $\gamma_{12} > \gamma_{23} > \gamma_{32} > \gamma_{21}$ , which was very close to the estimated results with noninformative prior (Table 5). However, artificial identifiability constraint may be inappropriate in our case as it is difficult to know the order of constraint with respect to 4 measurement errors if the pilot study is not conducted or other information is not available. As expected, the parsimonious model is therefore the one using the noninformative prior. Interestingly, the constraint order that is the same as that in the pilot study ( $\gamma_{23} > \gamma_{21} > \gamma_{32} > \gamma_{12}$ ) gives rise to the estimates of measurement errors (upper panel of Table 6) closer to our posterior estimates, except the fact that  $\gamma_{12}$  was a bit lower compared with other artificial constraints (Table 3). To further regard the parameters of measurement errors as hierarchical Dirichlet process (DP), we applied Bayesian hierarchical DP with conjugacy to model the hyperparameters of base measures in relation to 4 measurement errors given all 24 possible permutations with the iteration of Gibbs sampling scheme.<sup>31</sup> We think the results and statistical thoughts may be similar to those based on a probabilistic relabeling strategy although we have not used the same way as applied in the previous study.<sup>17</sup> It can be demonstrated that the estimated hyperparameters related to the coefficients of MetS using such a Bayesian hierarchical DP with conjugacy were close to those estimated in the main study with informative prior, which is clearly seen in Figure 5. However, the use of our proposed 2-stage Bayesian approach still gets the advantage. A statistically significant effect of MetS on the transition from moderate to severe AGA was

**TABLE 6** Estimated measurement errors, baseline transition rates, and regression coefficients of metabolic syndrome (MetS) adjusting for age, sex, and family history by using artificial identifiability constraints and Bayesian hierarchical Dirichlet process with conjugacy

	Normal to Intermediate			Intermediate to Severe		
	Estimate	SD	95% CI	Estimate	SD	95% CI
Artificial identifiability constraints approach with order the same as pilot study						
Proportion of misclassifications						
$\gamma_{12}$	9.29%	0.38%	8.49%-9.98%			
$\gamma_{21}$	9.64%	0.31%	9.03%-10.24%			
$\gamma_{23}$	13.99%	3.36%	9.70%-22.25%			
$\gamma_{32}$	9.49%	0.33%	8.83%-10.12%			
Baseline transition rates	0.0006	0.0001	0.0005-0.0007	0.0100	0.0020	0.0068-0.0148
MetS						
Yes vs no	0.0440	0.0416	-0.0392 to 0.1253	0.2044	0.0975	0.0051-0.3917
Age						
>70 vs $\leq$ 70	0.3255	0.0323	0.2630-0.3877	0.2076	0.0774	0.0578-0.3639
Sex						
Male vs female	2.4200	0.0904	2.2510-2.6020	-0.4751	0.1872	-0.8502 to -0.1156
Family history						
Yes vs no	1.4870	0.0694	1.3520-1.6250	0.6424	0.1277	0.3858-0.8936
Bayesian hierarchical Dirichlet process with conjugacy						
Proportion of misclassifications						
$\gamma_{12}$	27.08%	1.88%	23.38%-30.80%			
$\gamma_{21}$	7.82%	0.35%	7.11%-8.52%			
$\gamma_{23}$	15.46%	2.74%	10.48%-21.10%			
$\gamma_{32}$	9.90%	0.76%	8.43%-11.42%			
Baseline transition rates	0.0011	0.0001	0.0009-0.0013	0.0024	0.0008	0.0011-0.0042
MetS						
Yes vs no	0.0431	0.0428	-0.0389 to 0.1275	0.2197	0.1223	-0.0292 to 0.4479
Age						
>70 vs $\leq$ 70	0.3240	0.0342	0.2590-0.3940	0.3757	0.1090	0.1718-0.5997
Sex						
Male vs female	2.1080	0.0782	1.9560-2.2700	0.3800	0.2653	-0.1137 to 0.9516
Family history						
Yes vs no	1.4390	0.0842	1.2710-1.6000	1.0430	0.1598	0.7392-1.3670

noted whereas the corresponding point estimate based on the Bayesian hierarchical DP with conjugacy was close but gave a nonsignificant finding. This is entirely due to the incorporation of information from the pilot study that may increase statistical power. We believe such a 2-stage design may provide an alternative approach to tackling the problem of labeling switch problem commonly seen in the latent class and the mixed HMM. A formal statistical study in the future may be required to investigate the detailed comparisons between our proposed 2-stage design and the existing elegant statistical methods on probabilistic labeling switching algorithm.

There are one concern and one limitation of this study. Our proposed approach may be largely dependent on the performance of first-stage pilot study, especially when it was small. We performed a sensitivity analysis to explore the impact of the size of the pilot study (1/5, 1/2, 2, 5, 10, and 100 times the size of the current study) and the magnitude of measurement errors (half, double, and triple the measurement errors of the current study) in the pilot stage in our proposed Bayesian HMM. The results show that both ways may affect the estimated results of measurement errors only



**FIGURE 5** The scattered plot of regression coefficients of metabolic syndrome (MetS) on incidence and progression of androgenetic alopecia (AGA) by using different approaches

when the measurement error itself was large (data not shown). However, the small scale of the pilot study may fail to illustrate the statistical significance for the variable of main interests. In our AGA example, the size of the pilot study smaller than the current pilot study (100th of the main study) would lead to such a problem.

The major limitation that can be relaxed in a future study is that we merely investigated the measurement errors on outcome rather than on the explanatory factors, which may also be possibly subject to misclassification. From the viewpoint of application, because the covariate of major interests in the current study, MetS, was based on anthropometric measures (waist circumference and blood pressure) taken onsite by the trained public health nurses and biochemical examination, it was comparably objective. Except the self-reported data on family history, we reckoned the neglect of misclassification on covariates had limited influence on the application to AGA example. However, the methodology of simultaneously calibrating the measurement errors arising from covariates and also multistate outcome is still worthy of being developed.

In conclusion, we proposed a Bayesian measurement-error-driven HMM to deal with measurement errors of multistate outcomes specified by a continuous-time Markov process in order to calibrate the effect of covariates on the transitions of multiple states by making use of a 2-stage validation design. The simulation algorithms for assessing the direction of underestimation and overestimation were also developed to elucidate the underlying mechanisms accounting for 2 types (differential and nondifferential) and the possible range of underestimation and overestimation resulting from the misclassification of multistate outcomes. The proposed model has been applied to a real example of population-based follow-up study on multistate disease process of AGA.

## CONFLICT OF INTEREST

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## FUNDING

This study was supported by the Ministry of Science and Technology (grant number MOST 106-2118-M-002-006-MY2; MOST 106-2118-M-532-001-MY2; MOST 107-3017-F-002-003) and the Innovation and Policy Center for Population Health and Sustainable Environment (Population Health Research Center, PHRC) from Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan (grant number NTU-107L9003).

## ORCID

Hsiu-Hsi Chen  <http://orcid.org/0000-0002-5799-6705>

## REFERENCES

1. Hsieh HJ, Chen THH, Chang SH. Assessing chronic disease progression using non-homogenous exponential regression Markov models. *Stat Med.* 2002;21(22):3369-3382.

2. Yen AMF, Chen THH. Mixture multi-state Markov regression model. *J Appl Stat.* 2007;34(1):11-21.
3. Marshall G, Jones RH. Multi-state models and diabetic retinopathy. *Stat Med.* 1995;14(18):1975-1983.
4. Hubbard RA, Zhou XH. A comparison of non-homogeneous Markov regression models with application to Alzheimer's disease progression. *J Appl Stat.* 2011;38(10):2313-2326.
5. Pan SL, Wu HM, Yen AMF, Chen THH. A Markov regression random-effects model for remission of functional disability in patients following a first stroke: a Bayesian approach. *Stat Med.* 2007;26(29):5335-5353.
6. Wu YY, Yen MF, Yu CP, Chen HH. Individually tailored screening of breast cancer with genes, tumor phenotypes, clinical attributes and conventional risk factors. *Br J Cancer.* 2013;108(11):2241-2249.
7. Yen AMF, Wu WYY, Tabar L, Duffy SW, Smith RA, Chen HH. Initiators and promoters for the occurrence of screen-detected breast cancer and the progression to clinically-detected interval breast cancer. *J Epidemiol.* 2017;27(3):98-106.
8. Schork NJ. Personalized medicine: time for one-person trials. *Nature.* 2015;520(7459):609-611.
9. Zhang Y, Berhane K. Bayesian mixed hidden Markov models: a multi-level approach to modeling categorical outcomes with differential misclassification. *Stat Med.* 2014;33(8):1395-1408.
10. Dunson DB. Dynamic latent trait models for multidimensional longitudinal data. *J Am Stat Assoc.* 2003;98(463):555-563.
11. Miglioretti DL. Latent transition regression for mixed outcomes. *Biometrics.* 2003;59(3):710-720.
12. Altman RM. Mixed hidden Markov models: an extension of the hidden Markov model to the longitudinal data setting. *J Am Stat Assoc.* 2007;102(477):201-210.
13. Stephens M. Dealing with label switching in mixture models. *J R Statist Soc B.* 2000;62(4):795-809.
14. Jasra A, Holmes CC, Stephens DA. Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture model. *Stat Science.* 2005;20(1):50-67.
15. Chen THH, Duffy SW, Tabar L. A Markov chain method to estimate the tumour progression rate from preclinical to clinical phase, sensitivity and positive predictive value for mammography in breast cancer screening. *Stat.* 1996;45(3):307-317.
16. Jackson CH, Sharples LD, Thompson SG, Duffy SW, Couto E. Multistate Markov models for disease progression with classification error. *Stat.* 2003;52(2):193-209.
17. Sperrin M, Jaki T, Wit E. Probabilistic relabeling strategies for the label switching problem in Bayesian mixture models. *Stat Comput.* 2010;20(3):357-366.
18. Richardson S, Gilks WR. Conditional independence models for epidemiological studies with covariate measurement error. *Stat Med.* 1993;12(18):1703-1722.
19. Chen THH, Kuo HS, Yen MF, Lai MS, Tabar L, Duffy SW. Estimation of sojourn time in chronic disease screening without data on interval cases. *Biometrics.* 2000;56(1):167-172.
20. Duffy SW, Chen HH, Tabar L, Day NE. Estimation of mean sojourn time in breast cancer screening using a Markov chain model of both entry and exit from the preclinical detectable phase. *Stat Med.* 1995;14(14):1531-1543.
21. Kalbfleisch JD, Lawless JF. The analysis of panel data under a Markov assumption. *J Am Stat Assoc.* 1985;80(392):863-871.
22. Zucchini W, McDonald IL. *Hidden Markov models for time series: an introduction using R.* Taylor & Francis Group: Boca Raton; 2009.
23. Copeland KT, Checkoway H, McMichael AJ, Holbrook RH. Bias due to misclassification in the estimation of relative risk. *Am J Epidemiol.* 1977;105(5):488-495.
24. Norwood OT. Male pattern baldness: classification and incidence. *South Med J.* 1975;68(11):1359-1365.
25. Su LH, Chen THH. Association of androgenetic alopecia with metabolic syndrome in men: a community-based survey. *Br J Dermatol.* 2010;163(2):371-377.
26. Ludwig E. Classification of the types of androgenetic alopecia (common baldness) occurring in the female sex. *Br J Dermatol.* 1977;97(3):247-254.
27. Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. Executive summary of the third report of the National Cholesterol Education Program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III). *JAMA.* 2001;285(19):2486-2249.
28. Alberti KG, Zimmet P, Shaw J. Metabolic syndrome—a new world-wide definition. A consensus statement from the International Diabetes Federation. *Diabet Med.* 2006;23(5):469-480.
29. Lyles RH, Lin J. Sensitivity analysis for misclassification in logistic regression via likelihood methods and predictive value weighting. *Stat Med.* 2010;29(22):2297-2309.
30. Guihenneuc-Jouyaux C, Richardson S, Longini IM Jr. Modeling markers of disease progression by a hidden Markov process: application to characterizing CD4 cell decline. *Biometrics.* 2000;56(3):733-741.
31. Teh YW, Jordan MI, Beal MJ, Blei DM. Hierarchical Dirichlet processes. *J Am Stat Assoc.* 2006;101(476):1566-1581.

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Yen AM-F, Chen H-H. Bayesian measurement-error-driven hidden Markov regression model for calibrating the effect of covariates on multistate outcomes: Application to androgenetic alopecia. *Statistics in Medicine*. 2018;37:3125–3146. <https://doi.org/10.1002/sim.7813>