Published in final edited form as:

Stat Med. 2012 December 20; 31(29): . doi:10.1002/sim.5514.

PREDICTING LONGITUDINAL TRAJECTORIES OF HEALTH PROBABILITIES WITH RANDOM-EFFECTS MULTINOMIAL LOGIT REGRESSION

Xian Liu, Ph.D.^{1,2} and Charles C. Engel Jr., M.D., M.P.H.^{1,2}

¹Deployment Health Clinical Center, Walter Reed Army Medical Center, Washington, DC 20307-5001

²Department of Psychiatry, F. Edward Hebert School of Medicine, Uniformed Services University of the Health Sciences, Bethesda, Maryland 20814

Abstract

Researchers often encounter longitudinal health data characterized with three or more ordinal or nominal categories. Random-effects multinomial logit models are generally applied to account for potential lack of independence inherent in such clustered data. When parameter estimates are used to describe longitudinal processes, however, random effects, both between and within individuals, need to be retransformed for correctly predicting outcome probabilities. This study attempts to go beyond existing work by developing a retransformation method that derives longitudinal growth trajectories of unbiased health probabilities. Variances of the predicted probabilities are estimated by using the delta method. Additionally, the covariates' regression coefficients on the multinomial logit function, not substantively meaningful, are transformed to the conditional effects on the predicted probabilities. The empirical illustration uses the longitudinal data from the Asset and Health Dynamics among the Oldest Old (AHEAD). Our analysis compares three sets of the predicted probabilities of three health states at six time points, obtained from, respectively, the retransformation method, the BLUP, and the fixed-effects approach. The results demonstrate that neglect of retransforming random errors in the random-effects multinomial logit model results in severely biased longitudinal trajectories of health probabilities as well as overestimated effects of covariates on the probabilities.

Keywords

Growth trajectories; health status; longitudinal data; multinomial logit regression; prediction bias; random effects

INTRODUCTION

When modeling longitudinal growth trajectories of health data, researchers often encounter response outcomes characterized with three or more ordinal or nominal categories. Random-effects multinomial logit regression is generally used to account for potential lack of independence that exists in longitudinally clustered data [1–2]. As a basic feature underlying the generalized linear modeling, random errors in a random-effects multinomial logit regression are viewed as multivariate normal with mean 0 and variance-covariance matrix Σ . In using parameter estimates to describe longitudinal processes of health probabilities, the

usual application is to retransform the multinomial logit parameter estimates – the fixed effects and the between-persons random effects – to predict such probabilities without taking within-persons random errors into consideration. This specification perhaps arises from a widely held belief that because between-persons random effects reflect the individual differences in the multinomial response given specified model parameters, specification of the within-persons variability is unnecessary.

A closer look at this conventional predicting approach suggests some caution in accepting the rationale involved in this computation. For a random-effects multinomial logit model, both between-persons and within-persons random effects cannot be ignored from the retransformation process; otherwise predicted probabilities of a multinomial response at each time point may be misspecified, in turn resulting in prediction biases in outcome health probabilities. This is because, when one retransforms a transformed linear function to predict untransformed probabilities, the normality of random errors in the multinomial logit function must be retransformed as well to a lognormal distribution. Consequently, the expectation of the multiplicative random effects, after retransformation, is not unity but some quantity greater than one [3]. Even if the true values of model parameters are known, analytic results generated from a random-effects multinomial logit model cannot convert to unbiased estimates of health probabilities without considering all variance and covariance components.

This study attempts to go beyond existing work by developing a retransformation method that derives unbiased health probabilities longitudinally from the random-effects multinomial logit model. Specifically, we are focused on retransforming two components of random effects, between persons and within persons. The delta method is applied to compute standard errors of the predicted probabilities, based on the approximate variance-covariance matrix of a mean multinomial logit function. As regression coefficients in the multinomial logit model are difficult to interpret substantively [4–5], we develop conditional effects of a given covariate on a set of predicted probabilities for aiding interpretation of the effects on longitudinal growth trajectories of health data. Lastly, an empirical illustration is provided for displaying this method.

SPECIFICATION OF RANDOM-EFFECTS MULTINOMIAL LOGIT REGRESSION

Given normally distributed random errors in the multinomial logit model, exponentiation of those random errors yields a set of lognormally distributed disturbances, with means and variances well defined. Let Y_{ij} denote the value of an ordinal or a nominal categorical variable with $(K + 1)$ levels (for example, a health state) associated with person i at time j . Conventionally, adding scalar between-persons random effects to the fixed-effects multinomial logit model, the probability that $Y_{ij} = k$ ($k = 1, \dots, K$) for person i at time j is given by

$$P_{ijk} = \Pr(Y_{ij} = k | \mathbf{x}_{ij}) = \left[1 + \sum_{l=1}^K \exp(\mathbf{x}_{ij}' \beta_l + v_{il}) \right]^{-1} \exp(\mathbf{x}_{ij}' \beta_k + v_{ik}) \\ = \left[1 + \sum_{l=1}^K \exp(\mathbf{x}_{ij}' \beta_l) \exp(v_{il}) \right]^{-1} \exp(\mathbf{x}_{ij}' \beta_k) \exp(v_{ik}), \quad (1)$$

where $\mathbf{x}_{ij}$ is the $(M+1) \times 1$ covariate vector including intercept and, in the context of a longitudinal analysis, a time variable t_j where $j = 1, \dots, n_i$. Likewise, β_k is the $(M+1) \times 1$ vector of unknown regression parameters to be estimated, and v_{ik} is the between-persons

random effect assumed to be distributed as $N(0, \sigma_{v_k}^2)$. The probability $P_{ij(K+1)}$ is viewed as the reference or residual probability given the constraint that a set of choice probabilities must sum up to unity. Here, for illustrative simplicity and analytic convenience without loss of generality, we specify a random intercept multinomial logit model in Equation (1).

Equation (1) does not specify a term for within-persons random errors. The underlying rationale is that the specification of between-persons random effects reflects the individual differences in unspecified characteristics thereby addressing the within-persons variability [2, 6–7]. The best linear unbiased prediction (BLUP), widely applied in longitudinal data analysis, is primarily based on this specification [2]. In certain situations, Equation (1) does not necessarily reflect the true experiences generated by the stochastic longitudinal process because within-persons variability can have some unique impact on a multinomial response, even conditional on between-persons random effects [1].

If within-persons variability, assumed to be independent of $\mathbf{x}_{ij}$ and v_{ik} , is considered, then the specification of $P_{ij} = k$ ($k = 1, \dots, K$) for person i at time j becomes

$$\begin{aligned} P_{ijk} &= \Pr(Y_{ij} = k | \mathbf{x}_{ij}) = \left[1 + \sum_{l=1}^K \exp(\mathbf{x}'_{ij} \beta_l + v_{il} + \varepsilon_{ijl}) \right]^{-1} \exp(\mathbf{x}'_{ij} \beta_k + v_{ik} + \varepsilon_{ijk}) \\ &= \left[1 + \sum_{l=1}^K \exp(\mathbf{x}'_{ij} \beta_l) \exp(v_{il} + \varepsilon_{ijl}) \right]^{-1} \exp(\mathbf{x}'_{ij} \beta_k) \exp(v_{ik} + \varepsilon_{ijk}), \end{aligned} \quad (2)$$

where ε_{ijk} is the within-persons random disturbances distributed as $N(0, \sigma_{\varepsilon_k}^2)$.

Equation (2) can be conveniently transformed into a combination of linear specifications for K log odds. Let the $(K+1)$ 'th level be the reference category, then the inverse of Equation (2) for level k is

$$\log \left(\frac{P_{ijk}}{P_{ij(K+1)}} \right) = \mathbf{x}'_{ij} \beta_k + v_{ik} + \varepsilon_{ijk}, \quad \text{where } k=1, \dots, K. \quad (3)$$

Because an individual's probability is not empirically observable at a specific time point, the inherent within-persons variability cannot be directly captured in the observed multinomial response. The fixed-effects multinomial logit model assumes random errors to be completely reflected in fixed effects, thus evading the specification of random disturbances. Within a longitudinal setting, this specification is inescapable because there are two sources of random errors and one of them, the between-persons, is observable and estimable. Ignoring the existence of those random effects would be based on the hypothesis that subject-specific observations are conditionally independent, which, as discussed in much of the literature on longitudinal growth modeling, is not valid due to dependence in clustered data [1–3, 6–7].

Some researchers recommend the application of the latent variable approach [8–10] to estimate the within-persons random errors in the random-effects multinomial logit model [1]. That is, this random-effects component on each logit function is assumed to follow a standard logistic distribution with mean 0 and variance $\pi^2/3$. This standardized approach specifies a constant variance of within-persons random errors, regardless of the response type and the number of covariates considered in a particular regression model. It is argued that, when the between-persons random effects are specified, not that much variability

remains in a binary or a multinomial response (2). Additionally, this approach does not specify covariance between two related logit components, thereby overlooking the multivariate nature of the multinomial logit function. Indeed, the assumption of a continuous latent distribution is just a rough model requirement [11].

Statistically, within-persons random errors can be approximated by using the score equation, the first partial derivatives of the log likelihood in estimating a mean multinomial logit function. If all covariates are rescaled to be centered at some specified values, represented by $\mathbf{x}_0$, the logit intercepts correspond to a mean multinomial logit function with respect to $\mathbf{x}_0$, as applied in the regression analysis of correlated binary data [12] and for predicting the mean function of recurrent events in survival analysis [13]. For example, if time t is centered at 5 and other covariates are rescaled to be centered at sample means, the intercepts in the multinomial logit regression predict the mean logits for an average person at time 5. Correspondingly, the score function approximates the within-persons residuals for this mean multinomial function conditional on the between-persons random effects and other model parameters. Consequently, the variances and covariances of within-persons random errors can be estimated by the local subset of intercepts in the variance-covariance matrix for the fixed effects.

Given the approximates of $\sigma_{\varepsilon k}^2$, the probability that $Y_{ij} = k$ ($k = 1, \dots, K$) for person i can be predicted by retransforming the linear predictor in Equation (3), given by

$$\hat{P}_{ijk} = \left[1 + \sum_{l=1}^K \exp(\mathbf{x}'_{ij} \hat{\beta}_l) \hat{\Phi}_{il} \right]^{-1} \exp(\mathbf{x}'_{ij} \hat{\beta}_k) \hat{\Phi}_{ik}, \text{ for } k=1, \dots, K; l=1, \dots, K \quad (4)$$

where $\hat{\Phi}_{ik} = \exp(v_{ik} + \varepsilon_{ijk})$ is the estimated multiplicative random error variable for individual i on response category k ($k = 1, \dots, K$) given the definition of a lognormal distribution, with mean

$$E(\Phi_{ik}) = \exp\left(\frac{\sigma_{vk}^2 + \sigma_{\varepsilon k}^2}{2}\right),$$

and variance

$$\text{var}(\Phi_{ik}) = \exp\left[2\left(\sigma_{vk}^2 + \sigma_{\varepsilon k}^2\right) - \exp\left(\sigma_{vk}^2 + \sigma_{\varepsilon k}^2\right)\right].$$

Customarily, Equation (4) is defined as the inverse link function of the random-effects multinomial logit model, with the random components parameterized by two variance terms. The positive skewness of the lognormal distribution mandates that the expectation of Φ_{ik} is greater than unity, with equality holding if and only if $\sigma_{vk}^2 = \sigma_{\varepsilon k}^2 = 0$. Obviously, only using the fixed effects does not yield the unbiased estimate of a given probability marginalized at $\mathbf{x}_{ij}$, even if the true value of β is known. Conditional on β_k , the vectors of variances for the between-persons and the within-persons random errors for K categories, written as $\mathbf{V}(\mathbf{O})$ and $\mathbf{V}(\mathbf{E})$, respectively, contain values $\{\sigma_{v1}^2, \dots, \sigma_{vK}^2\}'$ and $\{\sigma_{\varepsilon 1}^2, \dots, \sigma_{\varepsilon K}^2\}'$, which may vary over an observation interval. The relative size of $\{\sigma_{v1}^2, \dots, \sigma_{vK}^2\}'$ determines whether the

within-persons random errors can be ignored in predicting the probabilities of $K+1$ response categories longitudinally.

If $\hat{\Phi}_{ik}$ and $\mathbf{x}_{ij}$ are replaced with $E(\Phi_{ik})$ and $\mathbf{x}_0$, respectively, Equation (4) predicts the marginal probability for a group of individuals with covariates valued at $\mathbf{x}_0$, with individual probabilities within the group randomly scattered around the marginalized probability according to the variability of both between-persons and within-persons random effects. This relationship between the subject-specific and the marginal logit model is graphically and analytically demonstrated in Diggle *et al.* [6] and Molenberghs and Verbeke [7]. Because the expected value of $\hat{\Phi}_{ik}$ is greater than 1, it is inappropriate to overlook this retransformed error term in specifying a marginal probability unless $\sigma_{vk}^2 = \sigma_{\varepsilon k}^2 = 0$ [7].

ESTIMATION OF MODEL PARAMETERS

Let θ be the vector of parameters with elements β and $V(O)$. Then, for individual i the likelihood function in the multinomial logit model can be written as a joint probability:

$$L(\mathbf{Y}_i|\theta) = \prod_{j=1}^{n_i} \prod_{k=1}^{K+1} (P_{ijk})^{Y_{ijk}}, \quad (5)$$

where Y_{ijk} is 1 if the i 'th observation falls in category k at time j and is 0 if otherwise, and $\mathbf{Y}_i$ is the $n_i \times (K+1)$ response matrix for individual i containing elements Y_{ijk} . While for $k=1, \dots, K$ P_{ijk} is defined by Equation (2), the estimation of P_{K+1} depends on the estimates of other probabilities in the same set, given by $\hat{P}_{K+1} = 1 - \hat{P}_1 - \dots - \hat{P}_K$.

Taking log values on both sides of Equation (5) gives rise to

$$\log L(\mathbf{Y}_i|\theta) = \sum_{j=1}^{n_i} \sum_{k=1}^{K+1} Y_{ijk} \log(P_{ijk}). \quad (6)$$

Maximizing the above function over all individuals requires that this log likelihood function be differentiated with respect to β and the random term $V(O)$. Given the specification of the between-persons random effects in θ , the estimation of parameter estimates on qualitative response data can be performed with integration over the random-effects. Let g be the logit link function and g^{-1} be its inverse function. Then, the expectation of P_{ijk} in Equation (4) is

$$E(P_{ijk}|\theta) = \int g^{-1} \left[\mathbf{x}'_{ij} \beta_k + \log(\Phi_{ik}) \right] dF(\Phi_i), \quad (7)$$

where the error distributional function F is the cumulative density function, and Φ_i is a vector of multiplicative random effects containing elements $\{\Phi_{i1}, \dots, \Phi_{iK}\}'$. In this integration, the differential term is written by $dF(\Phi_i)$ instead of by $dF(\Phi_{ik})$ because the prediction of the probability P_{ijk} involves all logit components. The term $V(E)$ is not particularly specified in θ because within-persons random errors are unobservable and considered to be embedded in the fixed effects [3]. If both σ_{vk}^2 and $\sigma_{\varepsilon k}^2$ are sizable, the predicted probability $\hat{P}_{ijk}$ can deviate markedly from the estimator $g^{-1}(\mathbf{x}'_{ij} \hat{\beta}_k)$. Because F is not a cumulative normal function, $\hat{P}_{ijk}$ is often not $g^{-1}(\mathbf{x}'_{ij} \hat{\beta}_k)$ thus highlighting the inevitability of retransforming random errors in the random-effects multinomial logit model. Whether $E(P_{ijk}|\theta) > g^{-1}(\mathbf{x}'_{ij} \hat{\beta}_k)$ or $E(P_{ijk}|\theta) < g^{-1}(\mathbf{x}'_{ij} \hat{\beta}_k)$ depends on the relative size of

Φ_{ik} in Φ_i . There are a variety of approximation approaches in this area proposed to derive Bayesian-typed estimators of θ given F [1, 14–18].

Let N be the total number of persons. Then, the maximum likelihood (ML) estimates of θ can be obtained by solving the following equation:

$$\frac{\partial \log L}{\partial \theta} = \sum_{i=1}^N g^{-1}(\mathbf{Y}_i) \left[\frac{\partial g(\mathbf{Y}_i)}{\partial \theta} \right] = 0. \quad (8)$$

As defined, the first partial derivatives of the log likelihood are the score function.

As a standard step of the Newton-Raphson procedure, the negative of the expected second partial derivatives of the log likelihood is the Fisher information matrix, denoted $I(\theta)$ and given by

$$I(\theta) = E \left(- \frac{\partial^2 \log L}{\partial \theta \partial \theta'} \right) = \sum_{i=1}^N g^{-2}(\mathbf{Y}_i) \frac{\partial g(\mathbf{Y}_i)}{\partial \theta} \left[\frac{\partial g(\mathbf{Y}_i)}{\partial \theta} \right]'. \quad (9)$$

According to the large-sample theory, the inverse of the observed information matrix approximates the variance-covariance matrix for parameter estimates in the multinomial logit model, from which within-persons variability can be approximated and evaluated by using the mean function approach. Hypothesis testing on linear combinations of the model parameters can be performed by calculating the generalized Wald statistic, distributed approximately as chi-square under the null hypothesis that $\beta = \mathbf{0}$ and $V(\mathbf{O}) = \mathbf{0}$ [1].

STANDARD ERRORS FOR PREDICTD PROBABILITIES

In predicting the longitudinal trajectories of health probabilities, standard errors of the predictions need to be computed for displaying and statistically assessing the dispersion of these estimates. We use the delta method for approximating such standard errors, as following the tradition in longitudinal growth modeling on nonlinear functions [2, 6–7, 19].

Let $\mathcal{L}$ be a random vector of the predicted multinomial logit function ($\mathcal{L} = \mathcal{L}_1, \mathcal{L}_2, \dots, \mathcal{L}_K$)' from Equation (3) with mean μ and variance-covariance matrix $\Sigma(\mathcal{L})$, and $\hat{\mathbf{P}} = g^{-1}(\mathcal{L})$ is a transform of $\mathcal{L}$, as defined by Equation (4), where g is the logit link function and g^{-1} is its inverse function. For large samples, the first-order Taylor series expansion of $g^{-1}(\mathcal{L})$ yields approximation of mean

$$E \left[g^{-1}(\hat{\mathcal{L}}) \right] \approx g^{-1}(\mu), \quad (10)$$

and the variance-covariance matrix $V(\hat{\mathbf{P}})$

$$V \left[g^{-1}(\hat{\mathcal{L}}) \right] \approx \left[\frac{\partial g^{-1}(\hat{\mathcal{L}})}{\partial \hat{\mathcal{L}}} \Big|_{\hat{\mathcal{L}}=\mu} \right]' \Sigma(\hat{\mathcal{L}}) \left[\frac{\partial g^{-1}(\hat{\mathcal{L}})}{\partial \hat{\mathcal{L}}} \Big|_{\hat{\mathcal{L}}=\mu} \right], \quad (11)$$

where

$$\frac{\partial g^{-1}(\widehat{\mathcal{L}})}{\partial \widehat{\mathcal{L}}} = \left[\frac{\partial g_1^{-1}(\widehat{\mathcal{L}})}{\partial \widehat{\mathcal{L}}}, \frac{\partial g_2^{-1}(\widehat{\mathcal{L}})}{\partial \widehat{\mathcal{L}}}, \dots \right],$$

and

$$\Sigma(\widehat{\mathcal{L}}) = \begin{bmatrix} \text{var}(\widehat{\mathcal{L}}_1) & \text{cov}(\widehat{\mathcal{L}}_1, \widehat{\mathcal{L}}_2) & \cdots & \text{cov}(\widehat{\mathcal{L}}_1, \widehat{\mathcal{L}}_k) \\ & \text{var}(\widehat{\mathcal{L}}_2) & \cdots & \text{cov}(\widehat{\mathcal{L}}_2, \widehat{\mathcal{L}}_k) \\ & & \ddots & \\ & & & \text{var}(\widehat{\mathcal{L}}_k) \end{bmatrix}.$$

In Equation (11), the matrix $V[g^{-1}(\widehat{\mathcal{L}})]$ is the approximate of the variance-covariance matrix $V(\mathbf{P})$ for large samples. The square-roots of the diagonal elements in this variance matrix yield standard errors of the predictions contained in the vector $\mathbf{P}$ given $\widehat{\boldsymbol{\beta}}$ and $V(\widehat{\boldsymbol{\theta}})$. As a result, the confidence interval for the predicted probability P_{ijk} can be easily calculated.

The delta method depends on the validity of the Taylor series approximation, so that some caution must be exercised before its adequacy is verified with simulation. As well justified in this analysis given $g^{-1}(\widehat{\mathcal{L}})$ to be a smooth nonlinear function of $\widehat{\mathcal{L}}$, the above procedure derives a robust, one-on-one standard error estimator [20]. Bootstrapping is sometimes applied for estimating standard errors of nonlinear predictions. The bootstrap procedure, however, generates less efficient estimates than does a retransformation approach with known elements in $\mathcal{Z}(\widehat{\mathcal{L}})$ for a multinomial logit model because it is based on the assumption that the off-diagonal elements in $\mathcal{Z}(\widehat{\mathcal{L}})$ are all 0 [21]. If elements in $\mathcal{Z}(\widehat{\mathcal{L}})$ are unobtainable empirically, some covariance structure in the multinomial distribution needs to be assumed for the application of the bootstrap.

Because an individual's probability is not empirically observable at a specific time point, the multinomial logit function does not have observed values, so the variance-covariance matrix for the predicted logits is not empirically obtainable. Here, a simple method is recommended to obtain an approximate of the matrix with K logit components. First, fit a multinomial logit model with all covariates rescaled to be centered at some specified values. Second, use the squared standard error of each intercept estimate plus the corresponding between-persons random intercept effect as the variance for each of the K logit components. Third, take values of covariance between each pair of the logit intercept estimates as the off-diagonal elements in $V(\widehat{\mathcal{L}})$. The rationale is that if covariates are centered at some specified values, the intercepts represent the group-specific means of logits, so that the local variance-covariance matrix for the estimated intercepts plus the corresponding between-persons random intercept effects can be considered the approximates of the variance/covariance matrix for the mean logit function. In the illustrative section, detailed procedures of this approximation method are described.

CONDITIONAL EFFECTS OF COVARIATES

One challenge about applying the multinomial logit model is interpretation of the estimated regression coefficients. Although displaying regression coefficients or odds ratios is not unpopular in interpreting results of the multinomial logit regression, such logit coefficients only provide a profile about the effects of covariates on the multinomial response. In the

presence of more than two response categories, regression coefficients in the multinomial logit model do not necessarily bear any relationship to changes in the probabilities themselves [4, 5, 22].

We propose to use the discrete probability change approach [10] for computing the conditional effects of covariates on the probabilities of a multinomial response. Consider the continuous covariate first. Let $\mathbf{x}$ be the vector containing sample means for all covariates, and $\mathbf{x}_r$ represents the vector containing sample means of the covariates other than a specific continuous variable x_m . Then, let $(P_k|\mathbf{x})$ be the marginalized estimate of $(P_k|\mathbf{x})$ when all covariates are scaled as sample means and let $(P_k|\mathbf{x}_m+1, \mathbf{x}_r)$ be another marginalized estimate of $(P_k|\mathbf{x})$ when x_m is scaled at one unit greater than its sample mean and other covariates are fixed at sample means. The difference between $(P_k|\mathbf{x}_m+1, \mathbf{x}_r)$ and $(P_k|\mathbf{x})$ is defined as the conditional effect of covariate x_m on the probability of category k marginalized at sample means, denoted by ΔP_{km} :

$$\Delta \hat{P}_{km} = \frac{\exp \left[\hat{\beta}_{km} (\bar{x}_m + 1) + \bar{\mathbf{x}}_r' \hat{\beta}_{kr} \right] \hat{\Phi}_k}{1 + \sum_{i=1}^K \exp \left[\hat{\beta}_{im} (\bar{x}_m + 1) + \bar{\mathbf{x}}_r' \hat{\beta}_{ir} \right] \hat{\Phi}_i} - \frac{\exp \left(\bar{\mathbf{x}}' \hat{\beta}_k \right) \hat{\Phi}_k}{1 + \sum_{l=1}^K \exp \left(\bar{\mathbf{x}}' \hat{\beta}_l \right) \hat{\Phi}_l}. \quad (12)$$

For a dichotomous variable, denoted x_m , the 0–1 change in that covariate should be specified. Accordingly, the conditional effect of x_m on the probability of category k is

$$\Delta \hat{P}_{km} = \frac{\exp \left(\hat{\beta}_{km} + \bar{\mathbf{x}}_r' \hat{\beta}_{kr} \right) \hat{\Phi}_k}{1 + \sum_{i=1}^K \exp \left(\hat{\beta}_{im} + \bar{\mathbf{x}}_r' \hat{\beta}_{ir} \right) \hat{\Phi}_i} - \frac{\exp \left(\bar{\mathbf{x}}_r' \hat{\beta}_{kr} \right) \hat{\Phi}_k}{1 + \sum_{l=1}^K \exp \left(\bar{\mathbf{x}}_r' \hat{\beta}_{lr} \right) \hat{\Phi}_l}. \quad (13)$$

Given the mathematical conditions that a set of probabilities always sum up to unity and the corresponding set of conditional effects must sum to zero, the conditional effect of a covariate on the probability of the $(K+1)$ 'th category is defined as $\Delta P_{K+1, x_m} = -(\Delta P_{x_j} + \dots + \Delta P_{Kx_j})$. With these restrictive conditions, all conditional effects of a given covariate on a set of probabilities should be considered statistically significant if any one of them is statistically significant.

This discrete probability change approach differs conceptually from the “marginal” effect that represents the slope of a probability function and actually has no bound in value [23]. The advantage of using the discrete conditional effect in the multinomial logit model resides in its consistence with the traditional interpretation of regression coefficients, especially when qualitative factors taking more than two values are used as covariates.

Notice that a statistically significant effect of a covariate on the logit scale may not necessarily translate into a significant effect on the probability, and therefore, the significance tests on these two different scales need to be performed separately. In particular, statistical significance of the conditional effect can be tested by using the Wald chi-square statistic, denoted by $\chi_{W,k}^2$ [24]. Let P_{k0} represent $(P_k|\mathbf{x})$ and P_{k1} stand for $(P_k|\mathbf{x}_m+1, \mathbf{x}_r)$. Then, the following equation is defined from the delta method:

$$\chi_{w,k}^2 \approx \frac{(\hat{P}_{k1} - \hat{P}_{k0})^2}{\hat{V}(\hat{P}_{k0}) + \hat{V}(\hat{P}_{k1}) - 2\text{cov}(\hat{P}_{k0}, \hat{P}_{k1})}, \quad k=1, \dots, K, \quad (14)$$

As predicted from the same parameter estimates with control variables fixed at sample means, $\hat{P}_{k0}$ and $\hat{P}_{k1}$ are subject to the same probability distribution. Therefore, the two random variables $\hat{P}_{k0}$ and $\hat{P}_{k1}$ should be nearly perfectly correlated. Consequently, the equation for calculating the Wald chi-square statistic reduces to

$$\chi_{w,k}^2 \approx \frac{(\hat{P}_{k1} - \hat{P}_{k0})^2}{\hat{V}(\hat{P}_{k0}) + \hat{V}(\hat{P}_{k1}) - 2\sqrt{\hat{V}(\hat{P}_{k0})\hat{V}(\hat{P}_{k1})}}. \quad (15)$$

This Wald statistic is distributed asymptotically as chi-square with 1 degree of freedom under the null hypothesis that $(\hat{P}_{k1} - \hat{P}_{k0}) = 0$. The evaluation of $\Delta\hat{P}_{K+1}$, the residual effect in the same probability set, depends on the testing results of the non-reference conditional effects. That is, if any non-reference conditional effect is significant, $\Delta\hat{P}_{K+1}$ is statistically significant; if all of the non-reference effects are statistically insignificant, $\Delta\hat{P}_{K+1}$ is not statistically significant.

ILLUSTRATION

Data used for this illustration come from the Survey of Asset and Health Dynamics among the Oldest Old (AHEAD), a nationally representative investigation of older Americans. This survey, conducted by the Institute for Social Research (ISR), University of Michigan, is funded by National Institute on Aging as a supplement to the Health and Retirement Study (HRS). At present, the Survey consists of nine waves of investigation. The Wave I survey was conducted between October 1993 and April 1994. Specifically, a sample of individuals age 70 years or older (born in 1923 or earlier) was identified throughout the HRS screening of an area probability sample of households in the nation. This procedure identified 9,473 households and 11,965 individuals in the target area range. AHEAD obtains detailed information on a number of domains, including demographics, health status, health care use, housing structure, disability, retirement plans, and health and life insurance. Survival information throughout the follow-up waves has been obtained by a link to the data of National Death Index (NDI). The present research uses data of the first six waves (1993, 1995, 1998, 2000, 2002 and 2004). Given the illustrative nature, we randomly select 2,000 older persons from the original AHEAD sample for the analysis described below.

The multinomial response variable is health state, measured by the difficulty level in performing activities of daily living (ADL), instrumental activities of daily living (IADL), and other types of functional activities [25]. An individual is defined as functionally disabled if he or she has any degree of difficulty in performing at least one of those activities. Death between two successive waves is viewed as a competing risk. As a result, at each follow-up time, three health states are identified, “not functionally disabled,” “functionally disabled,” and “dead in the time interval,” denoted by health states 1, 2, and 3. The probability of falling in a given health state is denoted by P_k , where $k = 1, 2, 3$. Given the focus on delineating longitudinal trajectories of health probabilities, the main explanatory variable of this study is time, scaled as the number of years elapsed since the time origin of the AHEAD survey ($t = 0, 2, 5, 7, 9, 11$). Gender is regarded as another main explanatory variable, measured as a dichotomous variable with women = 1 and men = 0,

because we want to compare longitudinal trajectories of health probabilities between older men and women. Age and educational attainment, both continuous variables, are used as controls, and in estimating the model parameters they are rescaled to be centered at sample means. An interaction term between time and gender was considered for capturing the possible convergence of the gender effect, but its inclusion did not contribute significantly to the model fit thus was removed.

For analytic convenience and illustrative simplicity without loss of generality, we apply the random intercept multinomial logit regression, so the effects of covariates on the multinomial logit function are assumed to be fixed. If needed, the random coefficient multinomial logit regression can be readily specified and estimated by extending the random intercept perspective. The SAS PROC NLMIXED procedure, given its tremendous flexibility in modeling nonlinear random-effects regressions, is used to compute both the fixed and random effects for the multinomial logit model [2, 7]. The adaptive Gaussian quadrature is applied for approximating the integral of the likelihood over the random effects given its advantage over the penalized quasi-likelihood (PQL) and the marginal quasi-likelihood (MQL) approximations in deriving robust estimates of random effects and the model fit statistic [7]. With the specification of between-persons random effects on intercepts, the variable “time” is treated as a continuous variable. After some preliminary data analysis, we found a combination of time and time $\times$ time, the quadratic polynomial time function, best describes the longitudinal trajectories of health probabilities in the AHEAD health data. Coming from the same source of variability, however, these two time components are highly correlated thus causing some numeric instability. As a result, the time variable t is rescaled to be centered at six (the midpoint of the observation interval) for substantially reducing numeric instability and multicollinearity.

To compare statistical efficiency and robustness of different methods for predicting longitudinal trajectories of health probabilities, we first create a “full” multinomial logit model including all covariates and two random effects components in the estimation, from which we assume the subject-specific and marginal probabilities and their variances/covariances to be exactly predicted. Then we purposefully exclude the variable “education” from the multinomial logit regression. As education has a statistically significant effect on the multinomial response in presence of other parameters, there is definitely additional clustering in the health data after its removal. Here, our operational objective is three-fold. First, we examine whether the new retransformation method can capture the random effects after removing an important predictor, as relative to the results obtained from the full model. Next, we want to know whether the retransformation method predicts longitudinal trajectories of health probabilities significantly better than does the BLUP. Lastly, we demonstrate how the fixed-effects perspective, though generating unbiased estimates of regression coefficients [6], results in severely biased predictions and the corresponding dispersions. The BLUP and the retransformation method are based on the same model that eliminates education from the estimation, and therefore, we develop three multinomial logit models in this illustration: the full random-effects model, the reduced random-effects model removing education from the regression, and the reduced fixed-effects model.

While the BLUP predicts probabilities for each observation [2], we obtain the group-wise BLUPs by creating a scoring dataset specifying the outcome variable as missing and the independent variables as zero, with covariates rescaled to be centered at selected values corresponding to a specific group. In the scoring data, the random effect per individual should be present, therefore the mean of the BLUPs with $y = \text{missing}$ and x 's = 0 generates the predicted probabilities for a group or an average person taking selected values of covariates.

Table 1 displays the analytic results of the three models. All regression coefficients, except the effect of gender on the second logit component in the two reduced models, are statistically significant at $\alpha = 0.05$. The regression coefficients of time for both logit components are positive, while those of time $\times$ time are all negative, which, combined, suggest a trend of deceleration in both the prevalence of disability and the mortality rate among older Americans. The between-persons random effect on the first logit component is only 0.001, not statistically significant, suggesting that time and age explain almost all variability in disablement of older persons. On the other hand, the between-persons random effect on the second logit component is 0.440, very strongly statistically significant. The intercepts, the regression coefficients, and the corresponding standard errors (not reported in Table 1) in the reduced fixed-effects model are very close to those derived from the two random-effects models. Such a similarity is not surprising given the statistical property that the asymptotic process $\sqrt{n}(\hat{\beta} - \beta_0)$, where β_0 is the true parameter vector, tends to converge to a normal vector with mean 0 and the covariance matrix $I^{-1}(\beta, t)$ for large samples. It does not mean, however, that the fixed-effects logit model yields robust and consistent estimators for predicting probabilities in the longitudinal setting, as will be shown below.

The probability of each health state at each time point and the corresponding variance can be estimated by applying Equations (10) and (11), respectively. Within the context of this illustration, we identify four partial derivatives of two predicted probabilities with respect to two logit components. For P_1 , we have

$$\begin{aligned}\frac{\partial \hat{P}_1}{\partial \mathcal{L}_1} &= \left\{ \left[1 + \sum_{k|\bar{t}=1}^2 \exp(\widehat{\mathcal{L}}_{k|\bar{t}}) \right]^2 \right\}^{-1} \exp(\widehat{\mathcal{L}}_1) [1 + \exp(\widehat{\mathcal{L}}_2)] = B_{11}, \\ \frac{\partial \hat{P}_1}{\partial \mathcal{L}_2} &= \left\{ \left[1 + \sum_{k|\bar{t}=1}^2 \exp(\widehat{\mathcal{L}}_{k|\bar{t}}) \right]^2 \right\}^{-1} [-\exp(\widehat{\mathcal{L}}_1) \exp(\widehat{\mathcal{L}}_2)] = B_{12}.\end{aligned}$$

Likewise, the corresponding partial derivatives for P_2 are

$$\begin{aligned}\frac{\partial \hat{P}_2}{\partial \mathcal{L}_1} &= \left\{ \left[1 + \sum_{k|\bar{t}=1}^2 \exp(\widehat{\mathcal{L}}_{k|\bar{t}}) \right]^2 \right\}^{-1} [-\exp(\widehat{\mathcal{L}}_1) \exp(\widehat{\mathcal{L}}_2)] = B_{21}, \\ \frac{\partial \hat{P}_2}{\partial \mathcal{L}_2} &= \left\{ \left[1 + \sum_{k|\bar{t}=1}^2 \exp(\widehat{\mathcal{L}}_{k|\bar{t}}) \right]^2 \right\}^{-1} \exp(\widehat{\mathcal{L}}_2) [1 + \exp(\widehat{\mathcal{L}}_1)] = B_{22}.\end{aligned}$$

Let

$$\left[\frac{\partial g^{-1}(\widehat{\mathcal{L}})}{\partial \widehat{\mathcal{L}}} \Big|_{\widehat{\mathcal{L}}=\mu} \right] = \mathbf{B},$$

where

$$B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}.$$

Then, the variance-covariance matrix for the predicted probabilities $\hat{P}_1$ and $\hat{P}_2$ can be approximated by the following sandwich estimator:

$$V(\hat{P}) \approx B \sum (\hat{\mathcal{L}} B',$$

where

$$\sum(\hat{\mathcal{L}}) = \begin{pmatrix} \sigma_{\bullet 1}^2 & \sigma_{12} \\ \sigma_{21} & \sigma_{\bullet 2}^2 \end{pmatrix}$$

In the retransformation method, the total variance is defined as $\sigma_{\bullet k}^2 = \sigma_{vk}^2 + \sigma_{ek}^2$, where $k = 1, 2$, consisting of two components of random effects, between-persons and within-persons.

While the first variance component σ_{vk}^2 is directly obtainable from the random-intercepts multinomial logit regression, the within-persons variances can be approximated by the first two diagonal elements of the variance/covariance matrix for the fixed effects on a specific mean multinomial logit function, as indicated earlier. In the BLUP, on the other hand, only σ_{vk}^2 is considered [2].

To highlight disparities in different predicting methods, we compute four sets of the predicted probabilities on the three health states at six time points, using results from the full model, the retransformation method, the BLUP, and the fixed-effects model, respectively. The retransformation method and the BLUP use the analytic results of the same multinomial logit model but are based on different perspectives for handling the inherent random effects. Table 2 presents these predicted probabilities and the corresponding standard errors at six time points.

In Table 2, the mortality rate is shown to increase exponentially over time at the early and the middle stages, and then the increase slows down near the end of the observation period, reflecting a “selection of the fittest” process. Corresponding to the increased mortality, the probability of disablement declines considerably over the follow-ups as a higher proportion of the disabled persons is expected to die compared to an earlier time interval. In view of the relatively small size of standard errors for the probabilities of disablement, all the predicted probabilities in Table 2 are considered statistically significant given the constraint that a set of predicted probabilities must sum up to unity. Overall, the retransformation method generates the closest sets of predictions, both the point estimates and the standard errors, to those derived from the full model, reflecting its high efficiency and coverage for handling random effects. The BLUP provides close predictions at the early waves; in the last three time points, however, the predicted probabilities deviate markedly from those derived from the first two models. Additionally, the BLUP, not accounting for the inherent within-persons variability, results in severely underestimated standard errors of the predictions. The fixed-effects model obviously generates the least reliable predictions given massive deviations from those obtained from the three random-effects models. More significantly, the fixed-effects approach severely underestimates the standard errors of the predicted probabilities, much more so than does the BLUP.

The effects of covariates on the predicted probabilities can be also severely biased if retransformation of random effects is overlooked. Table 3 displays four sets of gender's conditional effects on the probabilities of three health states at six time points. We use Equation (13) to calculate these effects as gender is dichotomous. Women are more likely to be functionally disabled but less likely to die, consistent with the general pattern. The retransformation method derives the closest sets of gender's conditional effects to those obtained from the full model, particularly the Wald chi-square statistics. The BLUP generates fairly close conditional effects in early times, but in later waves both the absolute values and the Wald statistics are noticeably underestimated. The fixed-effects approach underestimates the absolute values of gender's conditional effects and, at the same time, overestimates the Wald statistic to an exceptional extent due to neglect of internal variability in the predicted probabilities. It is worth mentioning that different methods can lead to conflicting results in the significance test on the conditional effects. When the fixed-effects approach is used, the conditional effects of gender are statistically significant throughout the entire observation period. On the other hand, the use of the BLUP or the retransformation method generates some insignificant effects along the longitudinal course.

To graphically display the statistical efficiency and coverage of different methods, Figure 1 plots three sets of longitudinal trajectories of disablement and the corresponding 95% confidence limits, with all covariates other than time fixed at sample means, derived from, respectively, the retransformation method, the BLUP, and the fixed-effects perspective. Each of these trajectory sets is compared to a solid growth line and a shaded area representing an unbiased 95% confidence region generated from the full model. In Panel A, which compares the predictions of disablement between the retransformation method and the full model, the two sets of curves almost coincide, thereby demonstrating the efficiency and coverage of the retransformation method after an important predictor is removed. The predicted growth curve in Panel B, derived from the BLUP, displays noticeable separations from the solid line in later times. More seriously, the two 95% confidence limit lines appear much contracted as relative to the shaded area, highlighting the limited capability of the BLUP in handling bias from random effects. Panel C displays the predicted growth curve generated from the fixed-effects model, in which tremendous and systematic bias can be easily identified. In later time points, those predicted probabilities are absurd: even the lower confidence limit of the probability appears much above the solid line.

Figure 2 plots three longitudinal trajectories of mortality and the corresponding 95% confidence limits, obtained from the retransformation method, the BLUP, and the fixed-effects approach. The solid line and the shaded area in each panel are the growth curve and the 95% confidence region generated from the full model, used as a standard for comparison. Panel A demonstrates how well the retransformation method predicts the trajectory of the old-age mortality and the corresponding confidence limits, after one theoretically important, statistically significant predictor is removed. As shown in Panel B, the BLUP predicts mortality fairly nicely in the early stage, but then it falls off systematically from the solid line, with the 95% confidence limits much underestimated. In Panel C, the predicted longitudinal growth curve of mortality is completely off base, with the two 95% confidence limits very narrowly scattered around.

DISCUSSION

As regression coefficients on the multinomial logit function are not interpretable substantively, a supplementary procedure is to use the fixed-effect estimates to predict marginalized probabilities. This application, however, entails serious prediction bias due to misspecification of the distribution of random errors. For a random-effects multinomial logit model, random effects cannot be simply overlooked when predicting the marginal

probabilities. If random disturbances in a random-effects multinomial logit model are truly normally distributed, the multivariate normality on the logit scale must convert to a multivariate lognormal distribution for correctly predicting the probabilities of a multinomial response. Consequently, the expected value of the random multiplier is not unity but some quantity greater than one, unless both between-persons and within-persons random errors are zero for all individuals. Additionally, without retransforming the distribution of random effects, the variance of the predicted probability will be much underestimated, in turn confounding the significance test on the prediction. The BLUP, widely used in longitudinal data analysis, accounts for between-persons variability in predicting probabilities; nevertheless, this approach overlooks within-persons random component that may exist inherently even with the specification of between-persons random effects.

In this article, we compare the predicted probabilities of three health states at six time points, obtained from three predicting methods – the retransformation method developed in this study, the BLUP, and the fixed-effects perspective – as relative to a prespecified full model. In particular, we examine how each of these methods behaves in prediction after an important predictor variable is removed from the estimation process. The results of our empirical illustration demonstrate that failure of retransforming random effects in the random-effects multinomial logit model results in severely biased longitudinal trajectories of health probabilities and sizable underestimation of the corresponding standard errors, even though the estimated regression coefficients tend to be unbiased. These biases, in turn, exert substantial influences on the significance test concerning the effects of covariates on the predicted probabilities. We display in our illustration that after statistically significant fixed effects of gender on the logit scale are converted into conditional effects on probabilities, the estimate of the gender effect has a much widened dispersion on the probability scale, and older men and women do not necessarily display statistically significant differences in health probabilities at all time points. This loss of statistical significance is thought to be because the fixed effects are population-averaged, so that the subject-specific variability is disregarded; once such random effects are considered in computing the probabilities, the inherent variability in the predictions is much increased in turn lowering the value of the chi-square statistic. The BLUP reduces some of such bias but its effect is shown to be limited. Our new method, retransforming both between-persons and within-persons variability simultaneously, increases the statistical efficiency and coverage in predicting longitudinal trajectories of health probabilities with random-effects multinomial logit regression.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by the National Institute on Aging (NIH/NIA Grant No.: R03AG20140-01). The authors thank three anonymous reviewers for valuable comments on an earlier draft; we also acknowledge the assistance of Phoebe McCutchan in generating the figures for this article.

References

1. Hedeker, D.; Gibbons, RD. Longitudinal Data Analysis. Wiley; Hoboken, NJ: 2006.
2. Littell, RC.; Milliken, GA.; Stroup, WW.; Wolfinger, RD.; Schabenberger, O. SAS for Mixed Models. 2. SAS Institute, Inc; Gary, NC: 2006.
3. Zeger SL, Liang KY, Albert PS. Models for longitudinal data: A generalized estimating equation approach. *Biometrics*. 1988; 44:1049–1060. [PubMed: 3233245]
4. Greene, WH. Econometric Analysis. 5. Prentice; New Jersey: 2003.

5. Liu X, Liang J, Muramatsu N, Sugisawa H. Transitions in functional status and active life expectancy among older people in Japan. *Journal of Gerontology: Social Sciences*. 1995; 50B:S383–S394.
6. Diggle, PJ.; Heagerty, PJ.; Liang, KY.; Zeger, SL. *Analysis of Longitudinal Data*. 2. Clarendon Press; Oxford: 2002.
7. Molenberghs, G.; Verbeke, G. *Models for Discrete Longitudinal Data*. Springer; New York: 2010.
8. Amemiya, T. *Advanced Econometrics*. Harvard University Press; Cambridge, MA: 1985.
9. Bock, RD. *Multivariate Statistical Methods in Behavioral Research*. McGraw-Hill; New York, NY: 1075.
10. Long, JS. *Regression Models for Categorical and Limited Dependent Variables*. Sage Publications; Thousand Oaks, CA: 1997.
11. McCullagh, P.; Nelder, JA. *Generalized Linear Models*. 2. Chapman and Hall; New York: 1989.
12. Zhao LP, Prentice RL. Correlated binary regression using a quadratic exponential model. *Biometrika*. 1990; 77:642–648.
13. Lin DY, Wei LJ, Yang I, Ying Z. Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society, Series B*. 2000; 62:711–730.
14. Breslow NR, Clayton DG. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*. 1993; 88:9–25.
15. Breslow NR, Lin X. Bias correction in generalized linear mixed models with a single component of dispersion. *Biometrika*. 1995; 82:81–91.
16. Goldstein, H.; Rasbash, J.; Plewis, I.; Draper, D.; Browne, W.; Yang, M.; Woodhouse, G.; Healy, M. *A User's Guide to MLwiN*. Institute of Education: University of London; 1998.
17. Stiratelli R, Laird N, Ware JH. Random-effects models for series observations with binary response. *Biometrics*. 1984; 40:961–971. [PubMed: 6534418]
18. Zeger SL, Karim MR. Generalized linear models with random effects: A Gibbs sampling approach. *Journal of the American Statistical Association*. 1991; 86:79–86.
19. SAS. *SAS/STAT 9.2: User's Guide*. SAS Institute Inc; Cary, NC: 2009.
20. Stuart, A.; Ord, JK. *Kendall's Advanced Theory of Statistics*. 6. Edward Arnold; New York: 1994.
21. Follmann D. Modelling transitional and joint marginal distributions in repeated categorical data. *Statistics in Medicine*. 1994; 13:467–477. [PubMed: 8023029]
22. Liao, TF. *Interpreting Probability Models: Logit, Probit, and Other Generalized Linear Models*. Sage Publications; Thousand Oaks, CA: 1994. (Series: Quantitative Applications in the Social Sciences 101)
23. Petersen T. A comment on presenting results from logit and probit models. *American Sociological Review*. 1985; 50:130–131.
24. Klein, JP.; Moeschberger, ML. *Survival Analysis: Techniques for Censored and Truncated Data*. 2. Springer; New York: 2003.
25. Liu X, Engel CC, Kang H, Cowan D. The effect of veteran status on mortality among older Americans and its pathways. *Population Research and Policy Review*. 2005; 24:573–592.

Figure 1.

Figure 2.

Table 1Results of three multinomial logit models on functional status In older Americans ($n = 2,000$)

Explanatory variable And effect type	Log(P_1/P_3)		Log(P_2/P_3)	
	Parameter est.	Standard error	Parameter est.	Standard error
Full random-effects model				
<u>Fixed Effects:</u>				
Intercept	5.000***	0.131	2.999***	0.141
Time (centered)	0.099***	0.013	0.306***	0.019
Time $\times$ time (centered)	-0.112***	0.005	-0.145***	0.006
Gender (centered)	0.599***	0.083	-0.699***	0.119
Age (centered)	0.086***	0.006	0.197***	0.010
Education (centered)	-0.100***	0.011	-0.102***	0.016
<u>Random Effects:</u>				
Intercept	0.001	0.058	0.440***	0.048
Reduced random-effects model				
<u>Fixed Effects:</u>				
Intercept	4.999***	0.127	2.999***	0.139
Time (centered)	0.101***	0.013	0.305***	0.017
Time $\times$ time (centered)	-0.121***	0.005	-0.135***	0.006
Gender (centered)	0.600***	0.080	-0.070	0.113
Age (centered)	0.088***	0.006	0.197***	0.009
<u>Random Effects:</u>				
Intercept	0.001	0.056	0.440***	0.053
Reduced fixed-effects model				
<u>Fixed Effects:</u>				
Intercept	4.999***	0.128	3.000***	0.138
Time (centered)	0.102***	0.014	0.301***	0.017
Time $\times$ time (centered)	-0.112***	0.005	-0.119***	0.006
Gender (centered)	0.600***	0.081	-0.030	0.107
Age (centered)	0.086***	0.006	0.197***	0.008

* $0.05 < p < 0.10$;** $0.01 < p < 0.05$;*** $p < 0.01$.

Table 2

Predicted probabilities of “disabled” and “dead” at six time points With standard errors: AHEAD longitudinal survey ($n = 2,000$)

Functional Status	Time point					
	T0 (1993)	T1 (1995)	T2 (1998)	T3 (2000)	T4 (2002)	T5 (2004)
	Predicted probability generated from the full model					
Disabled	0.646 (0.017)	0.948 (0.006)	0.950 (0.028)	0.851 (0.082)	0.677 (0.140)	0.544 (0.150)
Dead	–	0.005 (0.004)	0.044 (0.028)	0.144 (0.082)	0.311 (0.143)	0.412 (0.162)
	Predicted probability generated from the retransformation approach					
Disabled	0.635 (0.014)	0.942 (0.006)	0.939 (0.034)	0.850 (0.081)	0.698 (0.135)	0.563 (0.146)
Dead	–	0.006 (0.004)	0.053 (0.034)	0.143 (0.082)	0.290 (0.137)	0.386 (0.159)
	Predicted probability generated from BLUP					
Disabled	0.634 (0.000)	0.943 (0.001)	0.945 (0.011)	0.873 (0.030)	0.745 (0.050)	0.605 (0.056)
Dead	–	0.005 (0.001)	0.046 (0.012)	0.120 (0.030)	0.242 (0.051)	0.341 (0.062)
	Predicted probability generated from the fixed effects approach					
Disabled	0.622 (0.011)	0.909 (0.005)	0.884 (0.005)	0.826 (0.006)	0.775 (0.007)	0.748 (0.013)
Dead	–	0.041 (0.003)	0.108 (0.005)	0.167 (0.006)	0.211 (0.007)	0.184 (0.011)

Note: The predicted probability of “not disabled” and its standard error depend on the results for the probabilities of the other two health states, and therefore, it is not displayed in the table.

Conditional effects of gender on probabilities of two health states (female = 1) With Wald chi-square statistics in parentheses: AHEAD longitudinal survey ($n = 2,000$)

Table 3

Functional Status	Time point					
	T0 (1993)	T1 (1995)	T2 (1998)	T3 (2000)	T4 (2002)	T5 (2004)
	Conditional effect of gender generated from the full model					
Disabled	0.142 (972.462)	0.020 (63.197)	0.093 (5.307)	0.110 (8.917)	0.149 (17.520)	0.153 (29.670)
Dead	-	0.005 (3.158)	-0.089 (4.691)	-0.107 (8.065)	-0.144 (14.595)	-0.132 (12.454)
	Conditional effect of gender generated from the retransformation approach					
Disabled	0.142 (1027.312)	0.062 (12.202)	0.093 (5.343)	0.109 (8.876)	0.127 (15.986)	0.156 (26.087)
Dead	-	-0.034 (2.823)	-0.089 (4.665)	-0.107 (8.058)	-0.121 (12.657)	-0.133 (10.456)
	Conditional effect of gender generated from BLUP					
Disabled	0.142 (8420.194)	0.057 (10.461)	0.081 (4.164)	0.095 (6.882)	0.115 (13.947)	0.142 (28.254)
Dead	-	-0.027 (1.879)	-0.077 (3.554)	-0.092 (6.151)	-0.109 (10.756)	-0.115 (9.703)
	Conditional effect of gender generated from the fixed effects approach					
Disabled	0.141 (1201.061)	0.054 (357.738)	0.080 (522.340)	0.092 (603.781)	0.110 (776.900)	0.135 (763.141)
Dead	-	-0.024 (171.626)	-0.076 (459.845)	-0.088 (547.627)	-0.103 (650.768)	-0.105 (389.045)

Note: The conditional effects of gender on the probability of “not disabled” depend on the testing result on the conditional effects on the probabilities of the other two health states.