Predicting Longitudinal Trajectories of Health Probabilities with Random-effects Multinomial Logit Model

Xian Liu, Ph.D.

Charles C. Engel, MD, MPH

Research Objectives

1. Develop a new retransformation method for correctly predicting longitudinal trajectories of health probabilities

2. Provide an estimator for calculating standard errors of the predicted probabilities using the delta method

3. Demonstrate serious prediction biases in predicted probabilities without considering retransformation of random errors.

Research Significance

1. It helps scientists be aware of complexity in using random-effects multinomial logit model to describe longitudinal health data

2. It provides a new method to correctly predict longitudinal trajectories of health probabilities

3. It provides more accurate health outcome data for policy-makers and scientists

Hypotheses

1. Unbiased parameters cannot convert to unbiased estimates of health probabilities without retransforming random errors

2. Standard errors of predicted probabilities are severely underestimated if random effects are ignored

3. Both between-persons and within-person random errors can be estimated within the random-effects multinomial logit model

Model Specifications

1. The random-effects MNL model:

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

2. The inverse link function of the above:

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

Where

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

And

$$\mathrm{Var}(\Phi_k) = \exp\left[2\left(\sigma_{vk}^2 + \sigma_{\varepsilon k}^2\right) - \exp\left(\sigma_{vk}^2 + \sigma_{\varepsilon k}^2\right)\right].$$

Model Specifications (continued)

3. Standard error of $\hat{\mathrm{P}}$ (delta method):

Let $\hat{\mathbf{L}}$ be a random vector of the predicted multinomial logit function and $\hat{\mathbf{P}} = g^{-1}(\hat{\mathbf{L}})$ is a transform of $\hat{\mathbf{L}}$. Then

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

And

$$\mathbf{V}\left[g^{-1}(\hat{\mathbf{L}})\right] \approx \left[\frac{\partial g^{-1}(\hat{\mathbf{L}})}{\partial \hat{\mathbf{L}}}\middle| \hat{\mathbf{L}} = \boldsymbol{\mu}\right]' \boldsymbol{\Sigma}(\hat{\mathbf{L}}) \left[\frac{\partial g^{-1}(\hat{\mathbf{L}})}{\partial \hat{\mathbf{L}}}\middle| \hat{\mathbf{L}} = \boldsymbol{\mu}\right],$$

Where

$$\frac{\partial g^{-1}(\hat{\mathbf{L}})}{\partial \hat{\mathbf{L}}} = \left[\frac{\partial g_1^{-1}(\hat{\mathbf{L}})}{\partial \hat{\mathbf{L}}}, \frac{\partial g_2^{-1}(\hat{\mathbf{L}})}{\partial \hat{\mathbf{L}}}, \ldots\right].$$

Model Specifications (continued)

4. Conditional effects of covariate m ($\Delta\hat{P}_{km}$):

Let $(\hat{P}_{k0}|\bar{\mathbf{x}})$ and $(\hat{P}_{k1}|\bar{x}_m + 1, \bar{\mathbf{x}}_r)$ are two marginalized probabilities. Then

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

Significance test on $\Delta\hat{P}_{km}$ uses the Wald chi-square statistic:

$$\chi^2_{W,k} \approx \frac{\left(\hat{P}_{k1} - \hat{P}_{k0}\right)^2}{\hat{V}\left(\hat{P}_{k0}\right) + \hat{V}\left(\hat{P}_{k1}\right) - 2\hat{V}\left(\hat{P}_{k0}\right)\hat{V}\left(\hat{P}_{k1}\right)}.$$

Illustration

1. Data Source – The Survey of Asset and Health Dynamics among the Oldest Old (AHEAD), Wave I through Wave VI; 2,000 persons were randomly selected
2. Three health states – disabled, not disabled, dead – at five follow up time points
3. Covariates – Time, time × time, gender, time × gender, age, and education.
4. Random intercept MNL model using SAS PROC.GLIMMIX.

Table 1. Results of random-effects multinomial logit models on functional status In older Americans: AHEAD longitudinal survey (n = 2,000)

Explanatory variable And effect type	$\log(P_1/P_3)$ Parameter est.	$\log(P_1/P_3)$ Standard error	$\log(P_2/P_3)$ Parameter est.	$\log(P_2/P_3)$ Standard error
Fixed Effects:				
Intercept	4.940***	0.127	2.879***	0.154
Time (centered)	0.058***	0.013	0.411***	0.032
Time × time (centered)	-0.112***	0.005	-0.140***	0.006
Gender (centered)	0.535***	0.114	-0.345**	0.161
Time × gender (centered)	-0.015	0.021	0.004	0.034
Age (centered)	0.077***	0.006	0.226***	0.014
Education (centered)	-0.104***	0.012	-0.138**	0.018
Random Effects:				
Intercept	0.000	–	1.465***	0.434
Model Chi-square	7852.26			

* 0.05 < P < 0.10; ** 0.01 < P < 0.05; *** P < 0.01

Table 2. Predicted probabilities of three functional statuses at six time points With standard errors: AHEAD longitudinal survey (n = 2,000)

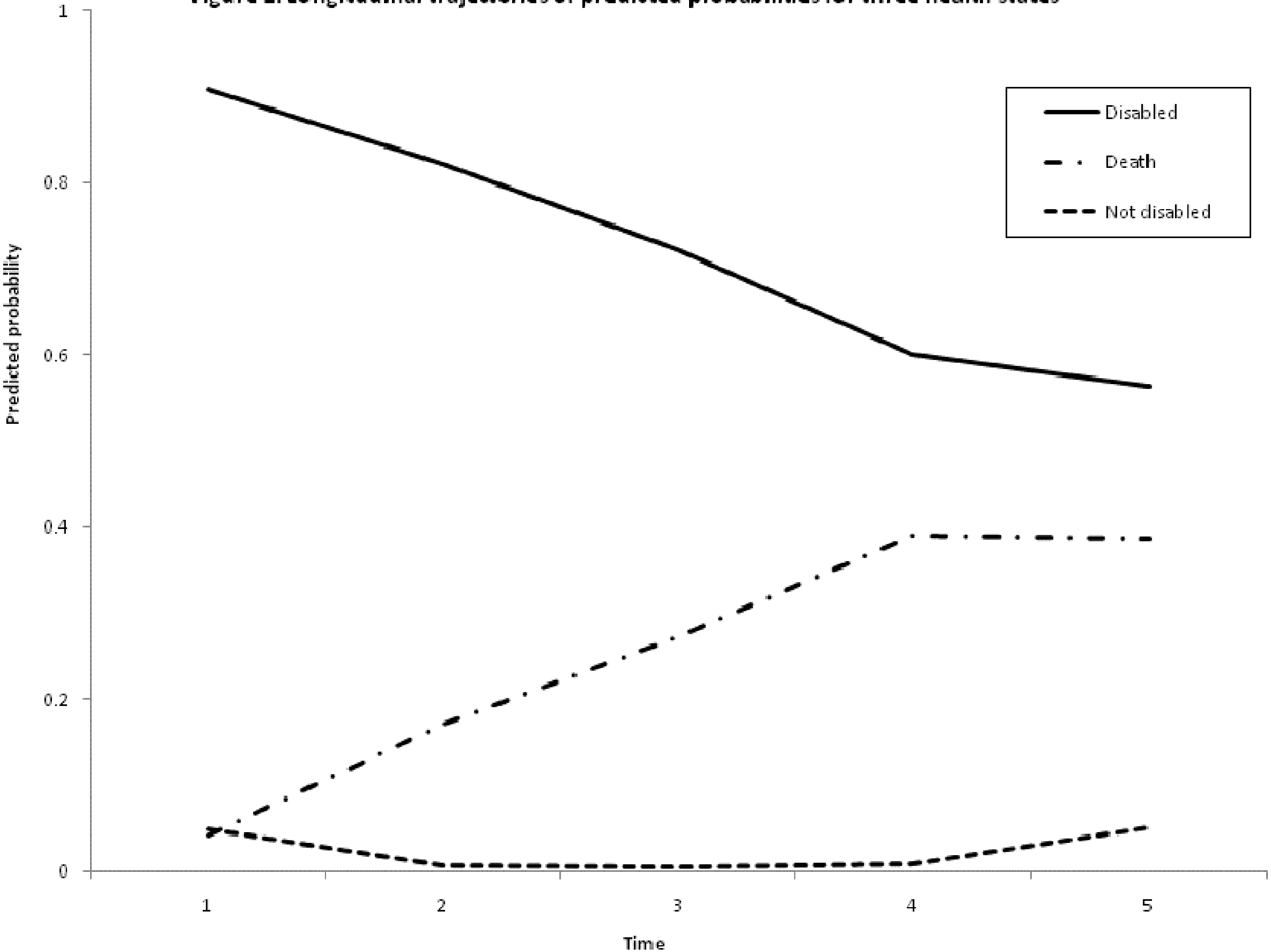
Functional Status	Time point					
	T0 (1993)	T1 (1995)	T2 (1998)	T3 (2000)	T4 (2002)	T5 (2004)
	Predicted probability generated from the retransformation approach					
Disabled	0.641	0.909	0.821	0.722	0.600	0.563
	(0.013)	(0.049)	(0.186)	(0.240)	(0.303)	(0.255)
Dead	–	0.043	0.172	0.272	0.390	0.387
		(0.051)	(0.188)	(0.241)	(0.308)	(0.278)
Not disabled	0.359	0.049	0.007	0.006	0.010	0.050
	Predicted probability generated from the fixed effects approach					
Disabled	0.641	0.930	0.912	0.843	0.770	0.696
	(0.011)	(0.005)	(0.008)	(0.009)	(0.010)	(0.018)
Dead	–	0.020	0.080	0.151	0.217	0.241
		(0.004)	(0.008)	(0.009)	(0.010)	(0.018)
Not disabled	0.359	0.050	0.008	0.007	0.013	0.063

Note: The test on the probability of "not disabled" depends on the testing results on the probabilities of the other two health states, and therefore, it does not have a standard error estimates

Table 3. Conditional effects of gender on probabilities of three functional statuses With chi-square statistics: AHEAD longitudinal survey (n = 2,000)

Functional Status	Time point					
	T0 (1993)	T1 (1995)	T2 (1998)	T3 (2000)	T4 (2002)	T5 (2004)
	Conditional effect of gender generated from the retransformation approach					
Disabled	0.149	0.070	0.141	0.192	0.198	0.164
	(35.702)	(0.651)	(0.227)	(0.263)	(0.233)	(0.222)
Dead	–	-0.042	-0.138	-0.191	-0.197	-0.150
		(0.206)	(0.214)	(0.256)	(0.223)	(0.156)
Not disabled	-0.149	-0.028	-0.003	-0.001	-0.001	-0.014
	Conditional effect of gender generated from the fixed effects approach					
Disabled	0.147	0.049	0.077	0.120	0.150	0.143
	(42.203)	(27.711)	(28.546)	(43.479)	(33.728)	(15.436)
Dead	–	-0.018	-0.073	-0.118	-0.147	-0.118
		(7.452)	(25.631)	(41.482)	(31.486)	(9.740)
Not disabled	-0.147	-0.030	-0.004	-0.002	-0.003	-0.024

Figure 1. Longitudinal trajectories of predicted probabilities for three health states



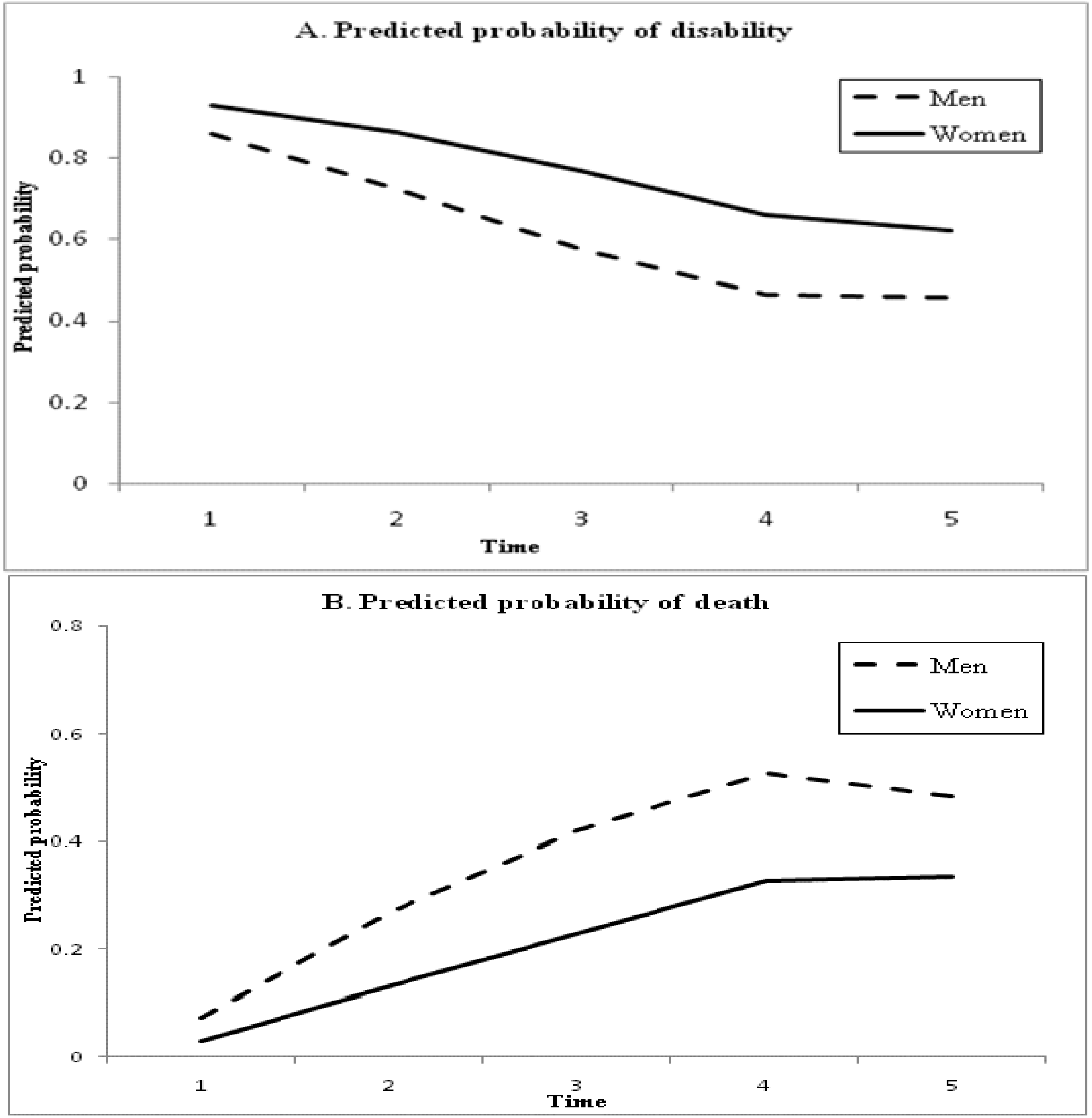


Figure 2. Predicted probabilities of disability and death by men and women

Conclusions

1. Neglect of random errors retransformation in the random-effects multinomial logit model leads to serious prediction biases in health probabilities.

2. Correspondingly, standard errors of those predicted probabilities are severely underestimated thereby resulting in misleading analytic results.