Joint analysis of growth curve model in depressive symptoms and mortality: application to an elderly study

Ping Yao
Correspondence: pyao@niu.edu
School of Nursing and Health Studies, Northern Illinois University, Dekalb, IL, USA.

Abstract
The primary aim in this study was to use a joint analysis approach to examine the association between longitudinal change of depressive symptoms and mortality in older Mexican Americans. The joint model was applied to data from the Hispanic Established Population for Epidemiological Study of the Elderly (HEPESE), a seven year longitudinal study of community-dwelling elderly Mexican-Americans. Depressive symptoms were measured by the Center of Epidemiological Studies Depression Scale (CES-D). The trajectories from two stage model and joint model using change in CES-D, modeled by longitudinal linear, and nonlinear curvatures including quadratic and exponential growth curve, were used as predictors to fit the survival curve. The joint analysis indicated that CES-D score at baseline and exponential curve were significant predictive factors of mortality and trajectories of linear and quadratic change in CES-D score were not associated with mortality in older Mexican-Americans.

Keywords: Longitudinal analysis, mexican americans, joint modeling, depression, mortality

Introduction
The prevalence of depression measured by the Center of Epidemiological Studies Depression Scale (CES-D) [1] in older Mexican-Americans is 25.4%, which is much higher than non-Hispanic Caucasians and African-Americans, whose prevalence ranges from 9-16.9% [2]. Because older adults in ethnic minorities including Mexican-Americans are exposed to higher level of depression risk factors, several studies have found higher levels of depressive symptoms in this group compared with non-Hispanic Whites [2]. The literature on associations between depression and mortality has been inconsistent. Depression negatively affects health and causes increasing mortality through complicated mechanisms [3], but the association of depression and mortality among the elderly population remains a controversy. A longitudinal community based study confirms an association of depression and mortality in Australia [4], and a cohort study finds that depression is a significant predictor for death in elderly people in Japan [5]. Several community based follow up studies reveal that mortality is associated with high baseline depressive symptoms [6-9]. On the other hand, there are reports showing that there is no association between depression and mortality [10-16].

Methods of joint modeling longitudinal repeated measurements and survival processes can be conducted to investigate these research questions. There are many methods developed to joint model repeated measures and time to event [17-25]. In general, a joint model is constructed from two sub-models, a repeated measurement model for longitudinal data and a survival model, which are linked by shared random effect. Depending on how to factorize joint distribution, these models can be classified into two categories: selection models and pattern mixture models. In the selection models, the linear mixed models characterize the repeated measurement and the censoring or missing mechanism is conditional on complete data [17-18]. In pattern mixture models, the samples are stratified into different subgroups by missing or censoring pattern, and then model the stratified subgroups separately [19-21] propose shared random-effect model. The repeated measures and death hazard are linked by shared random-effect.

Methods of jointly analyzing longitudinal and time to events can be conducted to investigate trajectory of depression and mortality [21-25]. With joint modeling, the longitudinal trajectory (intercept and slope) of the variable with repeated measures can be the predictors for the time to event, such as time to death. The joint model links two-sub functions, with one modeling repeated measures to capture change and the other survival function modeling time to death. Ghisletta et al., conducta joint analysis of longitudinal cognition and survival in an elderly population [26]. Zhang et al., use joint modeling of longitudinal change in depression symptoms to predict mortality ina community dwelling elderly population from Florida [13].

It’s very important to understand the progression and change trajectory of depressive symptoms and its effect on mortality risks in older Mexican-Americans. To our knowledge, it has not been addressed in previous studies. We examine the association of depression symptoms and mortality over seven years among older community dwelling Mexican Americans, from the Hispanic Established Population for the Epidemiological Study of the Elderly (HEPESE) [27-28].
This paper presents the application of joint analysis of growth curve model and time to event on interpretation of the trajectory effects of CES-D on death. In our HEPESE study, the objective is to use the longitudinal CES-D to predict the mortality of older Mexican-Americans. In Section 2, we introduce the samples and variables of HEPESE study. In Section 3, traditional separate models, two stage models, and joint models with linear, quadratic and nonlinear growth curve are applied to the HEPESE study. In Section 4, results are presented. We examine whether the baseline CES-D score and the rate of change of CES-D are predictors of mortality risk. Section 5 contains a discussion and conclusions.

**Study description**

**Data and sample**

We employed data from the HEPESE. This dataset comprised four waves of follow-up: 1993-1994, 1995-1996, 1998-1999 and 2000-2001. The sample consisted of 3050 Mexican-Americans at baseline, aged 65 years and older, residing in the five of the southwestern states of Arizona, California, Colorado, New Mexico, and Texas. Beginning with the second wave, some of the original subjects died or were lost to follow up and only 1683 subjects were interviewed at wave four [27,28].

**Measures**

All the adjusting independent variables were baseline values. The demographic variables included age, sex, years of education, annual household income, marital status, and body mass index (BMI). Marital status was dichotomized as unmarried (consisting of separated, divorced widowed and never married people) and currently married, with married people as the reference group. The number of chronic diseases was determined from seven items (cardiovascular disease, stroke, hypertension, diabetes, cancer, fractures and arthritis). The presence of disease was coded as “1” for “yes” or 0 for “no”. Possible scores for chronic diseases ranged from 0-7. The total number of limitations in activities of daily living (ADL), which included: walking across a small room, bathing, personal grooming, dressing, eating, getting from a bed to a chair, and using a toilet, was also an important physical functioning measurement at baseline. The range of this ADL scale was from 0-7, with higher scores indicating a greater number of ADL problems. Current smoking status (Smoking) and prescription drug use (Drug) with answer of yes or no. Possible scores for chronic diseases ranged from 0-7. The total number of limitations in activities of daily living (ADL), which included: walking across a small room, bathing, personal grooming, dressing, eating, getting from a bed to a chair, and using a toilet, was also an important physical functioning measurement at baseline. The range of this ADL scale was from 0-7, with higher scores indicating a greater number of ADL problems. Current smoking status (Smoking) and prescription drug use (Drug) with answer of yes or no at baseline was also used in this analysis. Cognitive status was assessed using Mini-Mental State Examination (MMSE).

**CES-D** The CES-D score consisted of the sum of 20 individual items measured on a four-point frequency scale, with score ranging from 0-60. Sample items include “I felt depressed,” “I felt fearful,” “I felt lonely,” and “I felt sad,” among others. Possible responses for each item on this 4-point Likert scale were: 0=never/none; 1=some/little; 2=occasionally/moderate; 3=most or all of time. Higher scores indicated greater level of depressive symptoms.

**Mortality** the time to death was calculated by the date of death minus the date of the first interview, measured by months.

**Statistical analysis**

**Model and estimation**

Suppose there are N subjects with repeated measure $Y_{ij}$ for the $i$th time point of subject $j$, where $j=1,2,\ldots,N$ and $i=1,2,\ldots,m$. Let $T_j=\min(D_j, C_j)$ be the observed survival time for the $j$th subject, where $D_j$ is the death time and $C_j$ is the censoring time, which is independent of $D_j$. Denote $\Delta_j$ the corresponding survival indicator, with $\Delta_j=1$ if $T_j=D_j$, and $\Delta_j=0$ otherwise.

First, descriptive statistics of dependent and independent variables at baseline were summarized. Next, Cox proportional hazard regression [29] was conducted to test whether the baseline CES-D score was predictive of mortality, with and without adjustment for baseline covariates. Kaplan-Meier Survival curve was plotted based on the dichotomized baseline CES-D score ($\geq 16$ indicating depression with indicator 1, and 0 otherwise).

Then the traditional two-stage growth curve model was conducted to estimate the individual change trajectory by ordinary least square estimation methods. At first stage, the linear regression was fitted to CES-D score against time for each individual subject. At the second stage, survival analysis was conducted by using individual linear intercept and slope of CES-D with adjusted by all covariates.

Finally, we conducted joint modeling of both longitudinal observations of CES-D and death intensity by shared random effect. The trajectory of change in CES-D scores was modeled by both linear and nonlinear growth curve. The intercept and slope of change in CES-D score were used as covariates in survival regression model to check the association of the trajectory of CES-D score over time with mortality, by adjusted models [13].

The model of repeated measure, linear growth curve of CES-D score was

$$Y_{ij} = \alpha_{0j} + \alpha_{1j} t_{ij} + e_{ij},$$

The nonlinear exponential growth curve of CES-D score was modeled by

$$Y_{ij} = \alpha_{0j} + \alpha_{1j} \exp(\tau t_{ij}) + e_{ij},$$

The nonlinear quadratic growth curve of CES-D score was

$$Y_{ij} = \alpha_{0j} + \alpha_{1j} t_{ij} + \alpha_{2j} t_{ij}^2 + e_{ij},$$

and denote $\mu_j = \alpha_{0j} + \alpha_{1j} t_{ij}$ for (1), $\mu_j = \alpha_{0j} + \alpha_{1j} \exp(\tau t_{ij})$ for (2), $\mu_j = \alpha_{0j} + \alpha_{1j} t_{ij} + \alpha_{2j} t_{ij}^2$ for (3), where $Y_{ij}$ was the CES-D score of subject $j$ at year $ij$; $e_{ij}$ was the residual, following a normal distribution with mean zero. $\tau$ was constant parameter. $\alpha_{0j}$, $\alpha_{1j}$ and $\alpha_{2j}$ were the subject-specific random intercept, slope of changes and nonlinear rate in the CES-D score of subject $j$; Assume $\alpha_{0j}$, $\alpha_{1j}$ and $\alpha_{2j}$ follow multivariate normal
distribution, which can be denoted by \( N(\alpha, \Sigma) \), i.e.,

\[
\begin{pmatrix}
\alpha_0j \\
\alpha_1j \\
\alpha_2j
\end{pmatrix}
\sim N(\alpha, \Sigma) = N
\begin{pmatrix}
\alpha_0 \\
\alpha_1 \\
\alpha_2
\end{pmatrix},
\begin{pmatrix}
\delta_0^2 & 0 & 0 \\
0 & \delta_1^2 & 0 \\
0 & 0 & \delta_2^2
\end{pmatrix}
\]

(4)

For (1) and (2), only \( \alpha_{0j} \) and \( \alpha_{1j} \) was from (4). Then the parameters of a linear or nonlinear mixed model were shared as covariates to predict mortality.

The Cox proportional hazards intensity function of the survival model for mortality was

\[
\lambda_j(t) = \lambda_0(t) \exp(\beta_0 + \alpha_{0j} \gamma_0 + \alpha_{1j} \gamma_1 + \beta_{chronic} + \beta_{age} + \beta_{sex} + \beta_{education} + \beta_{marital} + \beta_{income} + \beta_{ADL} + \beta_{MMSE} + \beta_{BMI} + \beta_{smoking} + \beta_{drug})
\]

The parametric Weibull hazards model at time \( t \) is given as

\[
\lambda_j(t) = \eta^{\gamma_j} \exp(\beta_0 + \alpha_{0j} \gamma_0 + \alpha_{1j} \gamma_1 + \beta_{chronic} + \beta_{age} + \beta_{sex} + \beta_{education} + \beta_{marital} + \beta_{income} + \beta_{ADL} + \beta_{MMSE} + \beta_{BMI} + \beta_{smoking} + \beta_{drug})
\]

The parametric Weibull hazards model shared quadratic term from (3) at time \( t \) is given as

\[
\lambda_j(t) = \eta^{\gamma_j} \exp(\beta_0 + \alpha_{0j} \gamma_0 + \alpha_{1j} \gamma_1 + \alpha_{2j} \gamma_2 + \beta_{chronic} + \beta_{age} + \beta_{sex} + \beta_{education} + \beta_{marital} + \beta_{income} + \beta_{ADL} + \beta_{MMSE} + \beta_{BMI} + \beta_{smoking} + \beta_{drug})
\]

Where \( \lambda_0(t) \) was the baseline hazard; \( \alpha_{0j} \) and \( \alpha_{1j} \) was shared from (1) and (2); the parameters \( \gamma_0, \gamma_1 \) and \( \gamma_2 \) were the subject-specific effects of the CES-D trajectory on mortality; \( \alpha_{2j} \) was shared from (3); \( \beta \) were corresponding parameters for control variables including chronic diseases, age, sex, education, marital status, income, ADL, smoking, cognitive status, drug and BMI.

Then likelihood of joint model for the \( j \)th subject is

\[
L_j = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2\sigma_i^2} \sum_{i=1}^{n} (y_{ij} - u_{ij})^2 \right) f(\alpha_j | \alpha, \Sigma) d\alpha_j
\]

where \( f(\alpha_j | \alpha, \Sigma) \) is the likelihood of the death of \( j \)th subject defined as

\[
L_j^D = \left[ \frac{\lambda_0(t)}{\lambda(t)} \right]^{\gamma_j} \exp\left( \beta_0 + \alpha_{0j} \gamma_0 + \alpha_{1j} \gamma_1 + \beta_{chronic} + \beta_{age} + \beta_{sex} + \beta_{education} + \beta_{marital} + \beta_{income} + \beta_{ADL} + \beta_{MMSE} + \beta_{BMI} + \beta_{smoking} + \beta_{drug} \right) dt
\]

The joint likelihood was maximized by quasi-Newton methods and Gaussian quadrature estimates with five quadrature points; the nuisance parameter, baseline hazard was approximated by a piecewise constant baseline hazard [22-25]. The analysis was done by the procedure NLMIXED in SAS 9.2 [30] and R package [31].

**Results**

(Table 1) illustrated the mean of CES-D score and the number of subjects who were followed, died or dropped out at each wave. The mean CES-D score were 14.91, 14.30, 15.46 and 6.68in the first, second, third and fourth waves. The CES-D score from wave four was not used since the value was not consistent with first three waves. The mortality rates were 7.84%, 21.67%, 30.59%, in the second, third and fourthwaves, respectively.

<table>
<thead>
<tr>
<th>Wave</th>
<th>N</th>
<th>Death</th>
<th>% (CES-D≥16)</th>
<th>Mean(SD) of CES-D</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>3050</td>
<td>0</td>
<td>36.71</td>
<td>14.91 (7.58)</td>
</tr>
<tr>
<td>II</td>
<td>2439</td>
<td>239</td>
<td>24.82</td>
<td>14.30 (5.85)</td>
</tr>
<tr>
<td>III</td>
<td>1981</td>
<td>661</td>
<td>33.92</td>
<td>15.46 (6.64)</td>
</tr>
<tr>
<td>IV</td>
<td>1683</td>
<td>933</td>
<td>13.73</td>
<td>6.68 (7.33)</td>
</tr>
</tbody>
</table>

(Table 2) presented the descriptive statistics of the study subjects at baseline stratified by overall, CES-D≥16 and CES-D<16.

(Figure 1) showed individual CES-D profiles of randomly selected subjects, which indicated nonlinear curvature of CES-D during seven year follow-up.

Kaplan-Meier Survival curve (Figure 2) indicated there was no significant difference between group of CES-D more than 16 with indicator 1 and group of CES-D less than 16 with 0 using baseline measurements.
First, Cox proportional regression was conducted to predict mortality using the baseline CES-D score. Without adjustment, baseline CES-D score was not significantly associated with mortality ($p=0.69$). After adjusting for age, sex, years of education, marital status, income, ADL scores, BMI, cognitive status, drug, smoke, and number of chronic diseases, through multivariate analysis, the baseline CES-D score was not a significant predictor of mortality ($p=0.68$).

The estimations of CES-D from two stage model showed the intercept with mean 14.73 (SD=7.15) and the slope with mean 0.05 (SD=3.35). The results of the second stage (survival analysis) indicated that the estimated baseline CES-D (intercept, $p=0.87$) and rate of change (slope, $p=0.95$) were not associated with mortality after adjusting by age, sex, years of education, marital status, income, BMI, cognitive status, drug, smoke, and number of chronic diseases. ADL was a very significant predictors of death ($p<0.01$).

Results of joint modeling of growth curve of CES-D and time to death

(Table 3) presented results of joint analysis of linear growth curve of CES-D and mortality adjusted by baseline variables to predict mortality at 7-year follow-up. The estimated intercept ($\alpha_0$) of CES-D score from the sub-model of the longitudinal part, which indicated the estimated mean baseline CES-D score, was 14.67 (SD=0.12, $p<0.001$). The linear slope ($\alpha_1$) of the trajectory of CES-D score was, 0.15 (SD=0.03, $p<0.001$), which was the estimated mean annual change in CES-D score. The results from multivariate analysis indicated the mean baseline of CES-D was significant in predicting mortality ($P<0.02$). The linear slope of CES-D was not a significant predictor ($P<0.17$). The effects of baseline covariates were also presented in (Table 3). At baseline, the results indicated that ADL was a very significant predictors of death. ADL score had a negative effect on mortality ($p<0.02$). Age, sex, years of education, marital status, income, BMI, cognitive status, drug, smoke, and number of chronic diseases at baseline were not significantly associated with mortality.

(Table 4) presented results of joint analysis of exponential growth curve of CES-D and mortality adjusted by baseline...
variables to predict mortality at 7-year follow-up. The nonlinear exponential model indicated the estimated intercept ($\alpha_0$) of CES-D score from the sub-model of the longitudinal part, which indicated the estimated mean baseline CES-D score, was 11.45 ($p<0.0001$). The exponential slope ($\alpha_1$) of the trajectory of CES-D score was, 3.18 ($p<0.0001$). The results from joint analysis showed the intercept of CES-D was significant in predicting mortality ($p<0.0001$). The exponential slope of CES-D was a significant predictor ($p<0.0001$). Both indicated the CES-D had negative effect on survival.

(Table 5) showed results of joint analysis of quadratic growth curve of CES-D and mortality. From (Table 5), we can see the estimated intercept ($\alpha_0$) of CES-D score was 5.43 ($p<0.0001$), slope was not significant, and quadratic curve ($\alpha_2$) of the trajectory of CES-D score was, 0.07 ($p<0.01$). The estimated intercept ($\alpha_0$) of CES-D score was significant associated with mortality ($P<0.0001$).

| Table 4. Results of joint analysis of exponential growth curve of CES-D and mortality adjusted by baseline variables to predict mortality at 7-year follow-up. |
|-----------------|-----------------|---------|-----------------|
| Predictor       | Estimate        | SE      | 95% CI           | P-value |
| CES-D intercept | -0.25           | 0.02    | -0.28 -0.22      | <.0001  |
| CES-D exponential| -0.50          | 0.03    | -0.56 -0.43      | <.0001  |
| Age             | -0.01           | 0.01    | -0.02 -0.00      | 0.20    |
| Sex (male)      | -0.09           | 0.09    | -0.27 -0.08      | 0.30    |
| Smoking (Yes)   | -0.09           | 0.12    | -0.34 -0.15      | 0.43    |
| Drug (Yes)      | 0.01            | 0.09    | -0.18 0.20       | 0.92    |
| Married (Yes)   | -0.05           | 0.09    | -0.23 -0.12      | 0.55    |
| Education       | -0.02           | 0.06    | -0.16 -0.10      | 0.68    |
| Income          | 0.02            | 0.03    | -0.04 0.08       | 0.48    |
| ADL             | -0.13           | 0.05    | -0.23 -0.02      | 0.01    |
| MMSE            | -0.004          | 0.01    | -0.02 0.02       | 0.74    |
| BMI             | 0.002           | 0.05    | -0.10 0.10       | 0.95    |
| Chronic         | 0.08            | 0.04    | 0.01 0.16        | 0.04    |

Estimate of exponential growth curve

$\alpha_0$: 5.43 0.72 4.01 6.84 <.0001
$\alpha_1$: -0.10 0.17 -0.45 0.23 0.54
$\alpha_2$: 0.07 0.02 0.02 0.13 0.01

Discussion

In this paper, we presented a joint random-effects model to analyze longitudinal depression symptom observations and death, to investigate the trajectory (intercept, slope and nonlinear growth curve) in depression symptom over time and the association with mortality in a seven year longitudinal study of older Mexican-Americans compared with the results from [13].

The results from Cox regression analysis using baseline covariates, indicated that baseline depression was not significantly associated with mortality ($p<0.67$). This result is consistent with dichotomized CES-D by Kaplan-Meier survival analysis.

Also, the results from joint longitudinal linear growth curve and survival analysis revealed that dynamic change of CES-D score was not a significant risk factor for mortality after controlling for all covariates in older Mexican-Americans. But results from joint longitudinal linear, exponential and quadratic growth curve and survival analysis showed that
the intercept (i.e., estimated baseline CES-D) was significant in predicting mortality, which is contradictory to what was found in the traditional Cox regression and Kaplan-Meier survival analysis.

Results of the current analysis are strengthened by the use of a large sample and long follow-up period to investigate the temporal association between depression and mortality. Furthermore, the sub-link of joint model, linear mixed model can capture the heterogeneity with random effects for individual subjects. The joint modeling can utilize the trajectory (intercept and slope) of depression to predict mortality in older Mexican-Americans.

In summary, the results presented from the joint analysis of longitudinal depression and mortality, indicated that the trajectory of change in CES-D score didn’t predict mortality in older Mexican-Americans. The intercept (baseline CES-score) is different from different analysis methods.

Competing interests
The author declares that he has no competing interests.

Acknowledgement
The authors thank Arlene Keddie from Northern Illinois University for providing the data of HEPES. The author are really appreciated the two referees’ comments so that the manuscript is improved.

References

Citation: