SHARED PARAMETER METHOD FOR MODELING THE EVOLUTION OF DEPRESSIVE SYMPTOMS IN LONGITUDINAL STUDIES WITH NONIGNORABLE MISSING DATA

by

HSIAO-CHING YANG

BSc, National Taiwan University, Taiwan, 1999

MSc, National Taiwan University, Taiwan, 2001

Submitted to the Graduate Faculty of

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2007

UNIVERSITY OF PITTSBURGH

Graduate School of Public Health

This thesis was presented

by

HSIAO-CHING YANG

It was defended on

July 24, 2007

and approved by

Vincent C. Arena, PhD, Associate Professor, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh

Mary Ganguli, MD, MPH, Professor, Departments of Psychiatry, Neurology, and Epidemiology, School of Medicine and Graduate School of Public Health, University of Pittsburgh

Thesis Director: Chung-Chou H. Chang, PhD, Research Assistant Professor, Departments of Medicine and Biostatistics, School of Medicine and Graduate School of Public Health, University of Pittsburgh

SHARED PARAMETER METHOD FOR MODELING THE EVOLUTION OF DEPRESSIVE SYMPTOMS IN LONGITUDINAL STUDIES WITH NONIGNORABLE MISSING DATA

Hsiao-Ching Yang, M.S.

University of Pittsburgh, 2007

In longitudinal studies of depressive symptoms in elderly patients, analyses are complicated by the presence of nonignorable missing data. In this study, we used data from the Monongahela Valley Independent Elders Survey (MoVIES) of 1,260 rural and elderly residents in western Pennsylvania. The method we used to analyze the evolution of depression is the shared parameter model, which is one of the methods that provide a framework for jointly modeling the longitudinal outcomes and the dropout process through a common frailty or unobserved random effects. When we used 2 different shared parameter models instead of using an unadjusted longitudinal model, we found the following decreases in the ratio of the odds of depression: a 2% decrease for women versus men (OR decreased from 2.05 in the unadjusted model to 2.00 in each shared parameter model); a 3% decrease for individuals with less than a high school education versus individuals with more than or equal to a high school education (OR decreased from 0.33 to 0.32); a 3% decrease for individuals taking fewer than 4 prescription drugs versus individuals taking 4 or more prescription drugs (OR decreased from 0.29 to 0.28); a 5% decrease for individuals using antidepressant drugs versus individuals not using antidepressant drugs (OR decreased from 16.15 to 15.35 in the first shared parameter model and to 15.39 in the second shared parameter model); and a 1% decrease for individuals with

functional impairment versus individuals without functional impairment (OR decreased from 4.72 to 4.66 in the first shared parameter model and to 4.67 in the second shared parameter model). Because differences of this magnitude are likely to have an impact on decisions concerning public health policies and funding, it is important to take nonignorable missing data into account when analyzing longitudinal studies. Shared parameter models can be computationally demanding, so their performance should be judged by their goodness of fit and required running time.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

In longitudinal studies of depressive symptoms in elderly patients, analyses are complicated by the presence of nonignorable missing data. Some patients are unwilling to participate further in scheduled follow-up interviews and examinations, some become too ill to do so, and some die before they are able to do so. Because severely ill or dying patients may in fact experience depression right before they drop out of the study, their missing data are nonignorable. If the problem of missing data is not handled appropriately, the study results may be biased and may consequently lead to inadequate management of depressive symptoms in the elderly population.

Missing data are usually categorized into three types (Rubin, 1976; Little and Rubin, 1987). For data missing completely at random (MCAR), the chance of missing does not depend on observed or unobserved values. For data missing at random (MAR), the chance of missing depends on observed but not unobserved values. For missing informative/nonignorable data (NI), the chance of missing may depend on unobserved values. The methods proposed to analyze studies with missing data include pattern mixture models and selection models. When the missing data type is NI, a joint model of the longitudinal outcome and missing process is often used.

In our study, the method we used to analyze the evolution of depression is the shared parameter model, which is one of the methods that provide a framework for jointly modeling the longitudinal outcomes and the dropout process through a common frailty or unobserved random effects (Wu and Carroll, 1988; Wulfsohn and Tsiatis, 1997; Lin et al., 2002; Roy, 2003; Tsiatis and Davidian, 2004; Beunckens et al, 2005; Vonesh et al., 2006). The key advantages of this approach are that it provides a flexible framework for handling nonmonotonic missing patterns and that it can be applied when study participants do not undergo the same number of follow-up interviews or examinations (follow-up "waves"). In our study, we modeled the evolution of depression by using scores from the modified Center for Epidemiologic Studies–Depression Scale (mCES-D Scale), and we modeled the time to dropout by using survival regression models of data collected from the Monongahela Valley Independent Elders Survey (MoVIES).

In Section 2 of this article, we introduce the notation and describe the shared parameter model that we used to analyze the evolution of depression. We also describe the procedure that we used to estimate the unknown parameters in the model. In Section 3, we begin by introducing the MoVIES data set and then present the analysis of results. In Section 4, we discuss possible extensions of our work.

2.0 METHODOLOGY: NOTATION AND MODEL

Let y_{ij} be a binary variable indicating whether individual i was depressed at wave j ($i = 1,..., N$ and $j = 1,..., J_i$). For individual i, let T_i be the time interval from baseline to dropout, let C_i be the observed censoring time (e.g., time interval from baseline to study end), and let $D_i = I(T_i < C_i)$ be the binary variable indicating whether the individual dropped out before the end of the study. Assume that D_{ij} is the binary variable indicating whether individual i dropped out before wave j. Let $\boldsymbol{X}_{ij}$ and $\boldsymbol{Z}_{ij}$ be the fixed-effect covariates associated with depression status and time to dropout, respectively. Note that $\boldsymbol{X}_{ij}$ and $\boldsymbol{Z}_{ij}$ may be overlapping. For example, age may be related both to depression status and to time to dropout, so the age variable will appear both in $\boldsymbol{X}_{ij}$ and $\boldsymbol{Z}_{ij}$. Let u_i be the shared parameter, which is the unobserved random effect contributing both to the probability of depression and the time to dropout. The association between longitudinal depression status and time to dropout is induced through this shared parameter.

The likelihood function of the shared parameter model we used to jointly model longitudinal depression status and time to dropout has the form

$$L = f(\boldsymbol{y}, \boldsymbol{T}) = \prod_{i=1}^{N} f(\boldsymbol{y}_i, T_i) = \int_{\boldsymbol{u}} f(\boldsymbol{y}, \boldsymbol{T} \mid \boldsymbol{u}) dF(\boldsymbol{u}) = \int_{\boldsymbol{u}} f(\boldsymbol{y} \mid \boldsymbol{u}) f(\boldsymbol{T} \mid \boldsymbol{u}) dF(\boldsymbol{u}), \tag{1}$$

where $f(\cdot)$ denotes a density function and $F(\cdot)$ denotes a cumulative density function. Note that under this model, if $f(\boldsymbol{T} \mid \boldsymbol{u}) = f(\boldsymbol{T})$, then $\boldsymbol{y}$ and $\boldsymbol{T}$ are independent. From this we infer that the dropout is ignorable.

The choice of $f(\boldsymbol{y} \mid \boldsymbol{u})$ depends on the type of longitudinal outcome. In our case, because the longitudinal response variable, depression status, is a binary variable, we used the binomial density function in $f(\boldsymbol{y} \mid \boldsymbol{u})$. We then used logistic mixed-effects regression to model the longitudinal outcome. This regression has the form

$$\operatorname{logit}\{\Pr(y_{ij} = 1) \mid \boldsymbol{X}_{ij}, u_i\} = \beta_0 + u_i + \boldsymbol{X}_{ij}\boldsymbol{\beta}, \ \ j \geq 2, \tag{2}$$

where $\boldsymbol{\beta}$ is a vector of fixed-effect parameters and the quantity $\Pr(y_{ij} = 1)$ can be interpreted as the depression prevalence. We assume that the shared parameter u_i follows a normal distribution with mean 0 and unknown variance σ^2.

There are many choices to model the time-to-event dropout in their studies. For example, for $f(\boldsymbol{T} \mid \boldsymbol{u})$, Hogan and Laird (1997) chose nonparametric models; Tsiatis et al. (1995) chose semiparametric models; Schluchter (1992), DeGruttola and Tu (1994), Pulkstenis et al. (1998), Schluchter et al. (2001), Guo (2004), and Vonesh et al. (2006) chose accelerated failure time models; and Vonesh et al. (2006) chose other parametric models and discrete failure time models. We chose the Weibull accelerated failure time model and the discrete failure time model.

The Weibull accelerated failure time model is useful because it is flexible enough to incorporate density functions with a wide range of shapes and because it often yields a reasonably robust estimate of γ, the scale parameter, provided that the assumption of proportional hazards is met (Vonesh, 2006). The hazard function of the Weibull model that corresponds to the component $f(\boldsymbol{T} \mid \boldsymbol{u})$ in (1) has the form

$$\lambda_i(t) = \lambda_0(t)\exp(\boldsymbol{Z}_i\boldsymbol{\alpha} + \alpha_{p+1}u_i), \tag{3}$$

where the baseline hazard function is $\lambda_0(t) = \gamma\exp(\alpha_0)t^{\gamma-1}$ with an unknown shape parameter $\gamma > 0$.

The discrete failure time model is useful in our study because the dropout time could be at any follow-up wave. The density function of this model has the form

$$\Pr(T_i = j \mid u_i) = (\lambda_{ik} \mid u_i)\prod_{k=1}^{j-1}\{1 - (\lambda_{ik} \mid u_i)\}, \tag{4}$$

where the discrete time hazard rate is $\lambda_{ik} \mid u_i = 1 - \exp\{-\exp(\alpha_{ik} + \boldsymbol{Z}_i\boldsymbol{\alpha} + \alpha_{p+1}u_i)\}$ and where α_{ik} defines the baseline conditional survival probability in the interval between wave $k-1$ and wave k—that is, at all time $t \in (t_{k-1}, t_k]$. We chose each time t_k as the follow-up wave.

To estimate the unknown parameters $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, σ, and γ, we used the maximum likelihood approach. This approach is needed to solve the score equation of the likelihood function shown in equation (1). Solving this equation involves high-dimensional integration. In general, there are three ways to approximate a high-dimensional integration numerically: to approximate the data using a pseudoresponse variable (Beal and Sheiner, 1982, 1988, Sheiner and Beal, 1985), to approximate the integral using either a nonadaptive or adaptive gaussian quadrature (Pinheiro and Bates, 1995), and to approximate the integrand using the Laplace method (Beal and Sheiner, 1992; Vonesh, 1996; Wolfinger and Lin, 1997; Raudenbush et al., 2000). Note that the method using a pseudoresponse variable may not perform well if the outcome variable is binary with few repeated measurements per individual (Verbeke and Molenberghs, 2000), as occurs in our study. A gaussian quadrature method is used to replace the integral by a weighted sum. The higher the order of the method, the more accurate the approximation is. The tradeoff of using a higher-order method is the computational intensity. The Laplace approximation of the integrand is an order one adaptive gaussian quadrature, so it is usually less accurate than a higher-order gaussian quadrature. In our study, we used and compared the Laplace approximation and adaptive gaussian quadrature, both of which are implemented in the SAS version 9.0 procedure NLMIXED (SAS Institute, Cary, NC).

To summarize the estimation of the unknown parameters $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, σ, and γ, we used the following steps:

1. Begin with an initial guess of the shared parameter from the generalized linear mixed model of the same form as equation (2).
2. Substitute the estimated shared parameter from step 1 into the approximated log-likelihood function of the joint model.
3. Maximize the approximated log-likelihood function from step 2 to obtain $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\alpha}}$, $\hat{\sigma}$, and $\hat{\gamma}$.

3.0 DISCUSSION: THE MONONGAHELA VALLEY INDEPENDENT ELDERS SURVEY

We used data from the Monongahela Valley Independent Elders Survey (MoVIES), which was conducted from 1987 to 2002 in two counties of southwestern Pennsylvania that were economically depressed and had lower education levels. Inclusion criteria for study participation included being at least 65 years old at the time of recruitment, being fluent in English, and having a sixth grade education or higher (Ganguli, 1993). The initial study cohort consisted of 1,681 participants, 1,422 of whom were randomly selected from voter registration lists in the study area and an additional 259 volunteers from the same area (Ganguli, 1993). Participants were assessed at study entry (wave 1) and reassessed in a series of approximately biennial data collection waves. Between waves 1 and 2, a total of 340 participants died, relocated, or dropped out of the study, leaving 1341 participants to be assessed at wave 2.

We collected data about depressive symptoms for the first time in wave 2 (1989-1991), which is considered the data baseline for our current study. We used the modified Center for Epidemiological Studies–Depression (mCES-D) scale, in which scores range from 0 to 20. For our study, we considered <5 to indicate not depressed and ≥5 to indicate depressed, based on the suggestion of Ganguli et al. (1995), Rovner and Ganguli (1998), and Ganguli et al. (2006).

Of the 1,341 participants in wave 2, we excluded 48 who had dementia before the depression assessment, 21 whose depression data were incomplete, and 12 whose data in the

Mini-Mental State Examination (MMSE) were incomplete. This left 1260 participants whose demographic and clinical characteristics of the sample have been described by Andreescu et al. (2007). Of the 1260 participants, 48 (38.10%) completed waves 2 and 6 but skipped one or more waves between. In this case, we imputed the missing mCES-D score by using the average of the scores derived before and after the missing wave.

Of the 1260 participants, a total of 669 (53.10%) dropped out before or at wave 6, with 171 (13.57%) of the dropouts occurring between waves 2 and 3, with 154 (12.22%) occurring between waves 3 and 4, with 164 (13.02%) occurring between waves 4 and 5, and with 180 (14.29%) occurring between waves 5 and 6 (Figure 1). When we used a chi-square test to compare the depression status of two groups—the 81(12.11%) of participants who dropped out before or at wave 6 (the dropout group) and the 46(7.78%) of participants who completed wave 6 (the nondropout group)—we found a significant difference in the percentage of depressed participants (21.10% vs. 16.60%, respectively; $P < 0.043$). As Figure 2 shows, there was a rise in the percentage of participants who were depressed before the dropouts occurred, and depression was the least common in the participants who completed the most waves.

As described earlier, we used shared parameter models to analyze data on depression evolution and dropout. For depression evolution, we used a logistic model with random intercepts representing the subject-specific baseline depression status. The model included the following covariates measured at baseline (wave 2): gender, age (65-74 years, 75-84 years, and ≥85 years), education level (less than high school or at least high school completion), number of prescription drugs used (<4 or ≥4), use of antidepressant drugs (yes or no), and functional impairment in instrumental activities of daily living (IADL, yes or no). To measure time to dropout, we defined the failure date as the actual date of dropout and defined the censoring date as the study end date (wave 6). We then used a Weibull accelerated failure time model and a discrete failure time model, both of which were conditional on the subject-specific random effect, which served as a subject-specific covariate. The results are shown in the first two columns of Table 1. Next, we compared the results from the shared parameter models (which jointly modeled depression evolution and dropout) with the results from a naïve generalized logistic random-effects model (which only modeled depression evolution).

For each participant, if the missing pattern is nonmontonic (may have missing values in between two nonmissing waves), we imputed these values by averaging the adjacent available outcome values. From all three models (the naïve model, shared parameter model with Weibull drop out, and the shared parameter model with discrete failure drop out), age is not a significant risk factor of being more depressed (mCESD ≥5). Being female, with less than high school education, having more than 4 prescription drugs used, using antidepressant drugs, and having functional impairment are positive risks factors of being more depressed. The odds ratios for each of the factor are very similar across the models. For the two shared parameter models, being older, being male, having more than 4 prescription drugs used, and having functional impairment are positively associated with time to drop out. Although we identified more 4 factors that are related to drop out, the longitudinal association between the risk factor and the depression status are similar with and without adjusting for drop out.

To check the model goodness of fit, we used two different statistics: -2 × log-likelihood and Akaike information criteria (AIC). A smaller value of the -2 × log-likelihood and a smaller value of the AIC value indicate a better model fit. The shared parameter model with Weibull drop out has slightly larger -2 × log-likelihood than the discrete failure time drop out does (-2 × log-likelihood = 5822.9 and 5402.4, respectively). The shared parameter model with Weibull drop out also has slightly larger AIC value than the shared parameter model with discrete failure time drop out does (AIC = 5868.9 and 5436.4, respectively). Although from the results of these two statistics, shared parameter model with discrete time failure time drop out is a better model than the shared parameter model with Weibull drop out, the CPU time used for the shared parameter model with discrete failure time drop out is much more (CPU time =1:13:57.45 versus 41:12.64).

Table 2 shows the percentage of depressed participants at each wave, stratified by the number of completed waves and by the covariates measured at baseline. We found that regardless of the number of waves that participants completed, depression was significantly less likely to occur in men than in women (5.02% vs. 9.42%; P <0.001); in participants aged 65-74 years than in those aged 75-84 years (5.99% vs. 10.59%; P <0.001); in participants who did complete high school than in those who did not (5.04% vs. 12.59%; P <0.001); in participants who used ≥4 prescription drugs than in those who used <4 drugs (6.02% vs. 16.10%; P <0.001); and in participant who did not use antidepressant drugs than in those who did (7.37% vs. 25.00%; P <0.001); and in participants who had functional impairment in IADL than in those who did not (3.89% vs. 13.51%; P value <0.001).

4.0 CONCLUSION

Missing observations are common in longitudinal studies with repeated data measurements and must be handled appropriately, especially when the missing data are of the nonignorable type. The shared parameter model, a newer method to handle this type of data, is a joint model of longitudinal outcomes and noninformative dropout. Using this method, we found that the longitudinal results were similar with and without adjusting for the noninformative dropout. The reasons for this finding are not clear. Although Tsiatis and Davidian (2004) suggested that the joint model of longitudinal and time-to-event data may not completely eliminate bias, they did not explore this issue. Further investigations are therefore needed to determine why and under what conditions the bias caused by noninformative dropout could be eliminated after adjusting for it in the shared parameter model framework.

The shared parameter model can be computationally demanding. For our work, when we used a server with 8 Xeon processors running at 2.66 GHz with 32 GB of RAM and a 4-disk striped RAID array, we spent about 1 hour of CPU time to run each model. When we used a personal computer with Pentium 4 processors running at 2.20 GHz with 1 GB of RAM, we spent about 4 hours of CPU time to run each model. If multiple shared parameter models are used to fit the data, the performance of the models should be judged by the model goodness-of-fit statistics (e.g., log-likelihood or AIC) and should also take the computational efficiency into account.

APPENDIX A

SAS NLMIXED PROCEDURE FOR SHARED PARAMETER MODEL WITH LOGISTIC RANDOM-EFFECTS DEPRESSION EVOLUTION AND WEIBULL TIME-TO-DROPOUT

```
proc nlmixed data=weibull start tech=DBLDOG maxiter=5000000
    maxfunc=500000;
 parms b0=3 b11=0 b12=0 b2=0 b3=3 b4=0 b5=0 b6=0 b71=0 b72=0 b73=0
       b74=0 a0=0 a11=0 a12=0 a2=0 a3=0 a4=0 a5=0 a6=0 a7=0
       gamma=2 psi=1;

 bi=(b0+u)+ b11*age7584+ b12*age85+ b2*female+ b3*higheduc+
    b4*rxfrqdum+ b5*deprs2+ b6*iadlc+b71*w3+b72*w4+b73*w5+b74*w6;
 ai=a0+ a11*age7584+ a12*age85+ a2*female+ a3*higheduc+ a4*rxfrqdum+
    a5*deprs2+ a6*iadlc+ a7*u;
 Pi = exp(bi)/(1+exp(bi));
 Li = exp(ai);
 Hi = gamma *(Li**gamma)*(t**(gamma-1));
 ll_Y= (1-ind)*(response1*log(Pi/(1-Pi))+log(1-Pi));
 ll_T= (ind)*(response1*log(Hi)-(Li*t)**gamma);
 model response1 ~ general(ll_Y+ll_T);
 random u ~ normal(0,psi) sub=subid;
run;
```

APPENDIX B

SAS NLMIXED PROCEDURE FOR SHARED PARAMETER MODEL WITH LOGISTIC RANDOM-EFFECTS DEPRESSION EVOLUTION AND DISCRETE FAILURE TIME ON DROPOUT

```
proc nlmixed data=discrete start tech=dbldog;
  parms b0=3 b11=0 b12=0 b2=0 b3=3 b4=0 b5=0 b6=0 b71=0 b72=0 b73=0
        b74=0 a1=-1 a2=-1 a3=-1 a4=-1
        L11=0 L12=0 L2=0 L3=0 L4=0 L5=0 L6=0 L7=0 psi=1;

  bi=(b0+u)+b11*age7584+b12*age85+b2*female+b3*higheduc+b4*rxfrqdum+b5*
  deprs2+b6*iadlc+b71*w3+b72*w4+b73*w5+b74*w6;
  Li=a1*w3+a2*w4+a3*w5+a4*w6+L11*age7584+L12*age85+L2*female+L3*highedu
  c+L4*rxfrqdum+L5*deprs2+L6*iadlc;
  P_i = exp(bi)/(1+exp(bi));
  H_i = 1-exp(-exp(Li+L7*u));
  ll_Y= (1-ind)*(response*log(P_i/(1-P_i))+log(1-P_i));
  ll_T= (ind)*(response*log(H_i/(1-H_i))+log(1-H_i));
  model response ~ general(ll_Y+ll_T);
  random u ~ normal(0,psi) sub=subid;
run;
```

APPENDIX C

SAS NLMIXED PROCEDURE FOR NAIVE MODEL WITH LOGISTIC RANDOM-EFFECTS DEPRESSION EVOLUTIONE

```
 proc nlmixed data=one start;
  parms b0=3 b11=0 b12=0 b2=0 b3=3 b4=0 b5=0 b6=0 b71=0 b72=0 b73=0
        b74=0 sigma=1;

  bi=(b0+u)+ b11*age7584+ b12*age85+ b2*female+ b3*higheduc+
      b4*rxfrqdum+ b5*deprs2+ b6*iadlc+ b71*w3+b72*w4+b73*w5+b74*w6;
  Pi = exp(bi)/(1+exp(bi));
  model response ~ binary(Pi);
  random u ~ normal(0,sigma**2) sub=subid;
run;
```

TABLES

Table 1. Results from the naïve model and the shared parameter models*

	Naïve Model		Weibull Model		Discrete Time Model	
Covariates	Estimated OR or HR⊥ (95% CI)	*P* value	Estimated OR or HR (95% CI)	*P* value	Estimated OR or HR (95% CI)	*P* value
Longitudinal model						
σ^2	18.84	<0.001	17.47	0.001	17.51	0.002
Age 75-84	1.04 (0.51, 2.10)	0.92	1.08 (0.54, 2.14)	0.83	1.08 (0.54, 2.14)	0.83
Age >85	0.63 (0.13, 3.07)	0.57	0.74 (0.16, 3.41)	0.63	0.68 (0.14, 3.18)	0.63
Female	2.05 (1.03, 4.09)	0.04	2.00 (1.02, 3.90)	0.04	2.00 (1.02, 3.91)	0.04
At least high school education	0.33 (0.17, 0.63)	<0.001	0.32 (0.17, 0.61)	<0.001	0.32 (0.17, 0.61)	<0.001
# of prescription drugs used	3.50 (1.65, 7.42)	<0.001	3.55 (1.70, 7.41)	<0.001	3.56 (171, 7.43)	<0.001
Use of antidepressant drug	16.15 (2.36, 110.46)	0.005	15.35 (2.35, 100.04)	0.004	15.39 (2.35, 100.69)	0.004
Functional impaired	4.72 (2.40, 9.30)	<0.001	4.66 (2.41, 9.03)	<0.001	4.67 (2.41, 9.04)	<0.001

Table 1 continued						
Dropout model						
Age 75-84	·	·	1.56 (1.38, 1.76)	<0.001	1.96 (1.63, 2.35)	<0.001
Age >85	·	·	2.4 (1.95, 3.08)	<0.001	3.79 (2.68, 5.36)	<0.001
Female	·	·	0.72 (0.65, 0.81)	<0.001	0.61 (0.52, 0.73)	<0.001
At least high school education	·	·	1.01 (0.90, 1.14)	0.82	1.00 (0.84, 1.19)	0.98
# of prescription drugs used	·	·	1.50 (1.33, 1.70)	<0.001	1.83 (1.52, 2.21)	<0.001
Use of antidepressant drug	·	·	1.23 (0.92, 1.63)	0.16	1.28 (0.83, 1.96)	0.27
Functional impaired	·	·	1.14 (1.01, 1.29)	0.03	1.22 (1.02, 1.46)	0.03
CPU (hr: min: sec)	4:56.85		41:12.64		1:13:57.45	
– 2×Log L(*y, T*)	1870.8		5822.9		5402.4	
AIC for L(*y, T*)	1896.8		5868.9		5436.4	

*All models are multivariable models with intercept and also adjusted for time (not shown)

⊥OR: odds ratio; HR: hazard ratio

Table 2. Percent of depressed individuals at each wave by individual's characteristics and by the individuals with different number of completed waves.

	Number of completed waves																			
	5 waves						4 waves					3 waves				2 waves			1 wave	
	N	2	3	4	5	6	N	2	3	4	5	N	2	3	4	N	2	3	N	2
Total	591	7.78	4.74	4.40	4.57	6.26	180	7.78	6.11	6.67	12.78	164	13.41	11.59	15.24	154	11.04	18.18	171	16.37
Age																				
65-74	446	7.40	4.26	3.59	3.14	4.93	88	6.82	4.55	5.68	12.50	68	13.24	10.29	13.24	59	5.08	15.25	86	13.95
75-84	140	9.29	6.43	6.43	7.86	9.29	88	9.09	7.95	6.82	12.50	83	13.25	12.05	16.87	77	15.58	20.78	65	16.92
85+	5	0.00	0.00	20.00	40.00	40.00	4	0.00	0.00	25.00	25.00	13	15.38	15.38	15.38	18	11.11	16.67	20	25.00
Gender																				
Male	195	4.62	2.56	2.05	2.56	4.10	71	5.63	4.23	5.63	11.27	73	5.48	2.74	9.59	61	6.56	11.48	94	11.70
Female	396	9.34	5.81	5.56	5.56	7.32	109	9.17	7.34	7.34	13.76	91	19.78	18.68	19.78	93	13.98	22.58	77	22.08
Education																				
<High school	195	13.33	9.74	8.72	8.72	9.74	72	8.33	5.56	6.94	18.06	74	17.57	18.92	25.68	74	17.57	22.97	75	17.33
≥High school	396	5.05	2.27	2.27	2.53	4.55	108	7.41	6.48	6.48	9.26	90	10.00	5.56	6.67	80	5.00	13.75	96	15.63
# of Rx medication																				
<4	521	7.87	4.61	4.03	4.03	5.76	138	5.07	2.90	5.07	10.14	117	10.26	7.69	9.40	102	6.86	13.73	108	7.41
≥4	70	7.14	5.71	7.14	8.57	10.00	42	16.67	16.67	11.90	20.46	47	21.28	21.28	29.79	52	19.23	26.92	63	31.75
Depressed at Wave 2																				
No	580	7.76	4.66	4.31	4.48	6.03	173	7.51	5.78	6.36	12.72	161	12.42	10.56	14.29	145	8.97	15.17	165	15.15
Yes	11	9.09	9.09	9.09	9.09	18.18	7	14.29	14.29	14.29	14.29	3	66.67	66.67	66.67	9	44.44	66.67	6	50.00

Table 2 continued																				
IADL																				
No	408	6.13	3.43	1.96	1.72	2.94	83	8.43	3.61	3.61	2.41	83	8.43	4.82	7.23	57	3.51	7.02	92	6.52
Yes	169	11.83	7.69	9.47	8.88	12.43	70	7.14	8.57	10.00	18.57	61	19.67	18.03	26.23	78	16.67	25.64	68	27.94

FIGURES

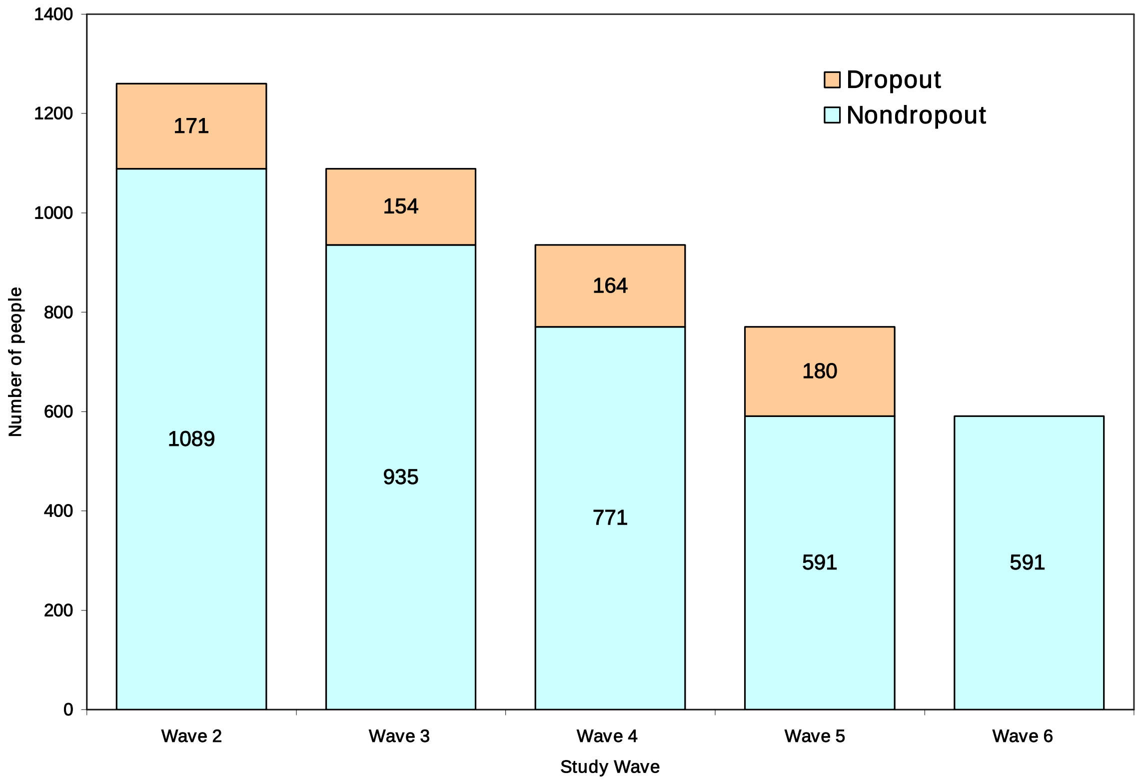

Figure 1. Dropout in each wave.

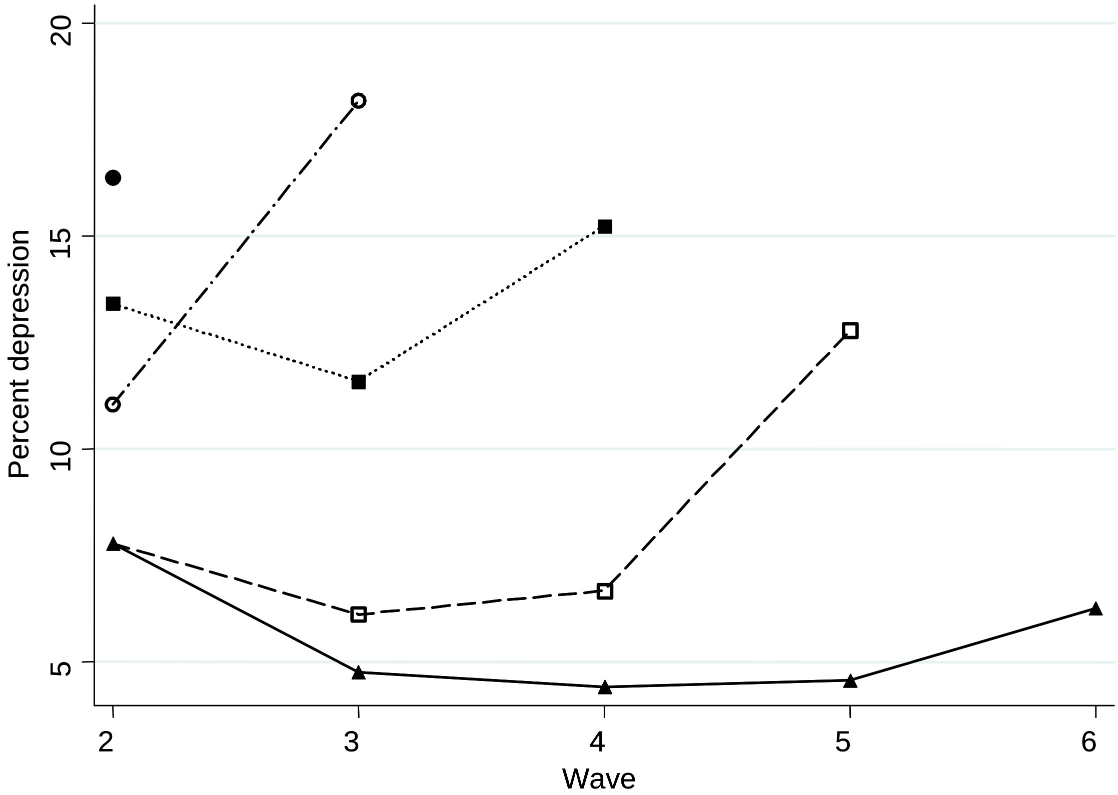

Figure 2. Percent depression by individuals who completed different number of waves: filled triangle and solid line (completed waves 2 to 6), unfilled square and dash line (completed waves 2 to 5), filled square and dot line (completed waves 2 to 4), unfilled circle and dash-dot line (completed waves 2 and 3), and filled circle (completed wave 2 only).

BIBLIOGRAPHY

Andreescu, C, Chang C. H, Mulsant, B, Ganguli, M. (2007). Twelve-year depressive symptom trajectories and their predictors in a community sample of older adults (Submitted).

Beal, S.L. and Sheiner, L.B. (1982). Estimating population kinetics, CRC Crit. Rev. Biomed. Eng., 8, 195–222.

Beal, S.L. and Sheiner, L.B. (1988). Heteroskedastic nonlinear regression, Technometrics, 30, 327–338.

Beal, S.L. and Sheiner, L.B., eds. (1992), NONMEM User's Guide, University of California, San Francisco, NONMEM Project Group.

Beunckens C, Molenberghs G, Kenward MG. (2005). Direct likelihood analysis versus simple forms of imputation for missing data in randomized clinical trials. Clin Trials, 2(5), 379-386.

De Gruttola, V. and Tu, X.M. (1994). Modelling progression of CD-4 lymphocyte count and its relationship to survival time. Biometrics, 50, 1003-1014.

Doble SE, Fisher AG. (1998). The dimensionality and validity of the Older Americans Resources and Services (OARS) Activities of Daily Living (ADL) Scale. J Outcome Meas, 2(1), 4-24.

Folstein, M., Folstein, S.E., McHugh, P.R. (1975). "Mini-Mental State" a Practical Method for Grading the Cognitive State of Patients for the Clinician. Journal of Psychiatric Research 1975; 12(3); 189-198.

Ganguli M, Belle S, Ratcliff G, Seaberg E, Huff FJ, von der Porten K, Kuller LH. (1993). Sensitivity and specificity for dementia of population-based criteria for cognitive impairment: the MoVIES project. J Gerontol, 48(4), M152-61.

Ganguli M, Dodge HH, Mulsant BH. (2002). Rates and predictors of mortality in an aging, rural, community-based cohort: the role of depression. Arch Gen Psychiatry, 59(11), 1046-52

Ganguli M, Gilby J, Seaberg E, Belle S. (1995). Depressive symptoms and associated factors in a rural elderly population: the MoVIES project. Am J Geriatr Psychiatry, 3, 144-160.

Ganguli M, Du Y, Dodge HH, Ratcliff GG, Chang CC. (2006). Depressive symptoms and cognitive decline in late life: a prospective epidemiological study. Arch Gen Psychiatry, 63(2),153-160.

Guo, X. and Carlin, B.P. (2004). Separate and joint modeling of longitudinal and event time data using standard computer packages. American Statistician, 58, 16-24.

Hogan, J.W. and Laird, N.M. (1997). Model-based approaches to analysing incomplete longitudinal and failure time data. Statistics in Medicine, 16, 259-272.

Lin, H., McCulloch, C.E. and Mayne, S.T. (2002). Maximum likelihood estimation in the joint analysis of time-to-event and multiple longitudinal variables. Statistics in Medicine, 21, 2369-2382.

Little R, Rubin D. (1987). Statistical analysis with missing data. New York: Wiley. Pulkstenis EP, Ten Have TR, Landis JR. (1998). Model for the analysis of binary longitudinal pain data subject to informative dropout through remedication. Journal of the American Statistical Association, 93, 438– 450.

Pinheiro, J.C. and Bates, D.M. (1995). Approximations to the Log-likelihood function in the nonlinear mixed-effects model, Journal of Computational and Graphical Statistics, 4, 12–35.

Raudenbush SW, Yang M-L, Yosef M. (2000). Maximum likelihood for generalized linear models with nested random effects via high-order multivariate Laplace approximation. Journal of Computational and Graphical Statistics, 9, 141–157.

Rovner B.W., Ganguli M. (1998). Depression and disability associated with impaired vision: the MoVies Project. Journal of the American Geriatrics Society, 46(5),617-9.

Roy, J. (2003). Modeling longitudinal data with nonignorable dropouts using a latent dropout class model. Biometrics, 59, 829-836.

Rubin, D.B. (1976). Inference and missing data. Biometrika, 63, 581–592.

SAS Institute Inc.: SAS OnlineDoc version 9, Cary, NC, SAS Institute Inc., 2006

Schlucter, M.D. (1992). Methods for the analysis of informatively censored longitudinal data. Statistics in Medicine, 11, 1861-1870.

Schluchter MD, Greene T, Beck GL. (2001). Analysis of change in the presence of informative censoring: application to a longitudinal clinical trial of progressive renal failure. Statistics in Medicine, 20, 989 –1007.

Sheiner, L.B. and Beal, S.L. (1985). Pharmacokinetic parameter estimates from several least squares procedures: Superiority of extended least squares, Journal of Pharmacokinetics and Biopharmaceutics, 13, 185–201.

Tsiatis, A.A. and Davidian, M.A. (2004). Joint modelling of longitudinal and time-to-event data: an overview. Statistica Sinica, 14, 809-834.

Tsiatis, A.A., DeGruttola, V. and Wulfsohn, M.S. (1995). Modelling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS. J. American Statistical Association, 90, 27-37.

Verbeke G, Molenberghs G. (2000). Linear Mixed Models for Longitudinal Data. New York: Springer-Verlag.

Vonesh, E.F., (1996). A Note on Laplace's Approximation in Nonlinear Mixed Effects Models. Biometrika, 83, 447–452.

Vonesh, E.F., Greene T. and Mark D. Schluchter M. D. (2006). Shared parameter models for the joint analysis of longitudinal data and event times. Statist. Med, 25, 143–163.

Wu, M. and Carroll, R. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. Biometrics, 44, 175-188.

Wolfinger, R.D., and Lin, X. (1997). "Two Taylor-series Approximation Methods for Nonlinear Mixed Models," Computational Statistics and Data Analysis, 25, 465–490.

Wulfsohn, M.S. and Tsiatis, A.A. (1997). A joint model for survival and longitudinal data measured with error. Biometrics, 53, 330-339.