*Iranian Int. J. Sci.* ***6****(1), 2005, p.85-105*

# Factors Affecting Age of Onset and Rate of Progression of MND

**E. Khoshbin[1] and R.B. Davies[2]**
[1]*Math. Stat. & Computer Sciences, Faculty of Science, University of Tehran, Tehran, Iran.*
[2]*Centre for Applied Statistics, Fylde College, Lancaster University, Lancaster, LA1 4YF.*
*(received: 3/5/2004 ; accepted: 1/11/2004)*

## Abstract

Background: As Motor Neuron Disease (MND) involves a progressive deterioration in nerve function, it is possible that risk factors for the disease will also accelerate the progression of the disease. Methods: This paper uses the two stages of age at onset and age at death to measure progression. Conventional parametric survival models are used to identify significant covariate from a large number of variables which have been suggested as risk factors for "MND / Frailty" models are used to characterize more fully the two stage progression process. Results and Conclusions: The results suggest the polluted work environments are associated with early onset of MND. Although there is a negative dependence between duration from onset to death and age at onset, this is not necessarily inconsistent with the expected progressive deterioration; there is a tantalizing suggestion from the frailty modeling that this may be the net result of a positive association due to variation in the rate of progression over the population and a negative association due to the direct effect of age (for example, because the elderly are more at risk from infections).

***Keywords:*** *MND, frailty models, true and spurious temporal dependence.*

## 1. Introduction

Motor neuron disease (MND), termed Amyotrophic Lateral Sclerosis (ALS) in the USA, is a fatal neurological disease. About one in 50,000 people worldwide develop MND each year (international Alliance of ALS/MND Associations) and approximately one in a thousand deaths in England and Wales are caused by this disease (Buckley *et al.*,1983).

However, there is limited understanding of the etiology of MND. In this paper we report an exploratory study using a rich dataset which includes environmental, family, and individual information for a sample of MND patients. We consider two stages of the disease: First, the onset of MND and then, second, eventual death from the disease. Studying these two stages has the obvious advantage of providing more insight into the progression of MND. It also enables us to model explicitly the effects of risk factors which have not been measured. It is well known that inferences about individual level temporal processes can be misleading unless the model accounts for the extra-variation between individuals due to omitted risk factors. This is achieved by incorporating a random effect in the model and such "frailty" models have received increasing attention in recent years (Lindley, D. *et al.,* 1986, Oakes D. 1989, Ling Mei. *et al.,* 1993, Dos Santos D. *et al.,* 1994, Hougaard P. 1995, Xue S. *et al.,* 1996, Yashin A. *et al.,* 1999). However, frailty modeling requires more than one duration or outcome for each individual if serious over-specification problems are to be avoided.

The majority of epidemiological analyses of MND have concentrated on two aspects:

- The Crude rise in mortality rate :

A crude rise in reported mortality rates from MND is evident in recent decades in a number of countries (Neilson S. *et al.,* 1993, Neilson S. *et al.,* 1994). One notable exception is mortality from MND in Japan, from 1950-1970 but the mortality rate has increased slowly subsequently (Neilson S. Robinson I. and Kondo K. 1993). One explanation suggested for this rise in mortality from MND is the increased life expectancy due to decrease in mortality from other causes to death. Another possible explanation is increasing exposure to environmental risk factors.

- Factors affecting MND :

Many studies have tested for associations between MND and environmental factors using aggregate level data and relationships have been claimed with employment in foundries, cotton mills, and coal mines, amongst others, (Neilson S. *et al.,* 1993, Mitchell J. *et al.,* 1995. Neilson *et al.,* 1992, Neilson *et al.,* 1993). Studies have also

reported various associations with illnesses in the past (for example, German measles, glandular fever, and chicken pox), and other individual and family characteristics. The research is reviewed by (Mitchell J. et al., 1995), and (Gatrell *et al.,* 1999).

Although motor neuron disease involves a progressive, irreversible deterioration in nerve function, we are not aware of any previous research relating age at onset to survival duration after onset. One possibility is that MND type deterioration is a characteristic of everyone and those suffering the disease are just those who by chance or action of one or more toxic agents have a particularly rapid rate of deterioration (Mitchell J. et al., 1999). If this is the case, then onset at a relatively young age would be associated with particularly extreme rates of deterioration and would imply short post-onset survival, with the reverse for onset in later year. We investigate this possibility in this paper.

In section 2 we describe the dataset used in this study. Initial data analysis and results of the explanatory modeling using conventional models are reported in section 3. Section 4 deals with joint modeling of the two durations, and section 5 concludes the study.

## 2. The Data

The data used in this study were collected at the neurology department of the Royal Preston Hospital which covers most of Lancashire and South Cambria.The data consist of detailed life history data for the 128 patients diagnosed with MND between Jan. 1980 and Feb. 1995. The lengthy data collection schedule was completed by a nursing research officer and covered family background, immunization history, travel abroad, previous heath, education and employment, diet, animal contacts, and places of residence. The questions were designed to cover most of the variables suggested in the literature to be associated with MND. Previous analyses of the data for these patients and control groups have suggested that there is some evidence of an environmental association, and MND appears to be less prevalent amongst the first born and only child in the family. It was found also that chickenpox was significantly related to survival time after onset. Clearly, the study design was concerned with identifying MND risk factors rather than studying duration to onset and death.

However, (Mitchell J. et. al 1995) conclude that some risk factors are associated with rate of progression as well as occurrence of the disease and this inspired the use of the dataset in this study.

The variables extracted from the dataset are listed in Table (1) together with percentages. These explanatory variables have all been suggested as risk factors for MND. It is noted that a number of the second durations (from onset to death) are right-censored, because the patient was still alive at the end of the study. Some categories of the explanatory variables have very low percentages. Where this is the "Don't know" category it is a welcome feature. However, the low percentages for Glandular Fever and some occupational exposures limit the inference possible.

**Table 1: Explanatory variables included in the dataset.**

| Variable | Level | % |
|---|---|---|
| **Demographic :** | | |
| Sex | 1 Male | 60% |
| | 2 Female | 39% |
| Siblings | 1 Had sibling (s) | 7% |
| | 2 Only child | 92% |
| **Medical History :** | | |
| German Measles | 1 Had German Measles | 21% |
| | 2 Never had German Measles | 57% |
| | 3 Don't know | 21% |
| Chicken Pox | 1 Had Chicken Pox | 46% |
| | 2 Never Had Chicken Pox | 32% |
| | 3 Don't know | 20% |
| Thyroid Disorder | 1 Had Thyroid Disorder | 11% |
| | 2 Never Had Thyroid Disorder | 88% |
| | 3 Don't know | 1% |
| Glandular Fever | 1 Had Glandular Fever | 2% |
| | 2 Never Had Glandular Fever | 96% |
| | 3 Don't know | 1% |
| **Occupational Exposures :** | | |
| Foundry | 1 Worked in Foundry | 12% |
| | 2 Never Worked in Foundry | 87% |
| Cotton Mill | 1 Worked in Cotton Mill | 14% |
| | 2 Never Worked in Cotton Mill | 85% |
| Coal Mine | 1 Worked in Coal Mine | 6% |
| | 2 Never Worked in Coal Mine | 93% |
| Quarry | 1 Worked in Quarry | 4% |
| | 2 Never Worked in Quarry | 96% |
| Agricultural Chemicals | 1 Worked with Agricultural Chem. | 11% |
| | 2 Never Worked with Agricultural Chem. | |
| | 3 Don't know | 2% |

**Table 2: Continuation of Table (1).**

| Variable | Level | % |
|---|---|---|
| Chemicals and Solvents Exposure | 1 Exposure | 39% |
| Industrial | 2 No Exposure | 58% |
| | 3 Don't know | 2% |
| Chemicals and Solvents Exposure | 1 Exposure | 14% |
| Laboratory | 2 No Exposure | 82% |
| | 3 Don't know | 3% |
| Fumes or Dust | 1 Exposure | 65% |
| | 2 No Exposure | 32% |
| | 3 Don't know | 1% |
| Lead | 1 Exposure | 15% |
| | 2 No Exposure | 82% |
| | 3 Don't know | 3% |
| Mercury | 1 Exposure | 3% |
| | 2 No Exposure | 94% |
| | 3 Don't know | 3% |
| Manganese | 1 Exposure | 4% |
| | 2 No Exposure | 90% |
| | 3 Don't know | 6% |
| Aluminum | 1 Exposure | 12% |
| | 2 No Exposure | 83% |
| | 3 Don't know | 5% |
| Sewage | 1 Exposure | 4% |
| | 2 No Exposure | 95% |
| | 3 Don't know | 1% |
| Rubber | 1 Exposure | 9% |
| | 2 No Exposure | 91% |
| | 3 Don't know | 1% |
| Any other Metals or Elements | 1 Exposure | 28% |
| | 2 No Exposure | 71% |
| | 3 Don't know | 1% |
| Contact with Cat | 1 Exposure | 26% |
| | 2 No Exposure | 50% |
| | 3 Don't know | 24% |

## 3. Initial Data Analyses

### *3.1 Kaplan Meier Estimates*

Kaplan Meier estimates for durations 1 (age at onset) and 2 (duration from onset to death) are shown in Figure 1. These plots show that the earliest recorded onset is at age 30 and that the majority of cases experience onset in the age range 60-80, with the median at age 62.2. The maximum recorded time from onset to death is 11.2 and the median is at 2.6 years. Complementary log-log transformations of these estimates are plotted in Figure 2. The evident curvature for duration from onset to death suggests that a Weibull model would not be appropriate. However, for age of onset the estimates are nearly

linear after the first few points on the distribution. This suggests that a Weibull model may be appropriate for age of onset.

### *3.2 Initial Modeling without Explanatory Variables*

Tables (3), (4), and (5) show the results of fitting a number of standard 2-parameter survival models for the two durations of interest and the age of death, again with no explanatory variables. The software package GLIM (Francis B. et al., 1993) was used to fit these models. The Weibull provides a better fit than the Gompertz for age at onset. Consistent with the Kaplan Meier plots, the log-logistic and log-normal models, which cannot represent monotonically increasing hazards, fit less well. The results are similar for age of death. For duration from onset to death, the situation is very different; the log-normal gives the best fit, marginally better than the log-logistic, while the Weibull and Gompertz are decidedly worse fitting.

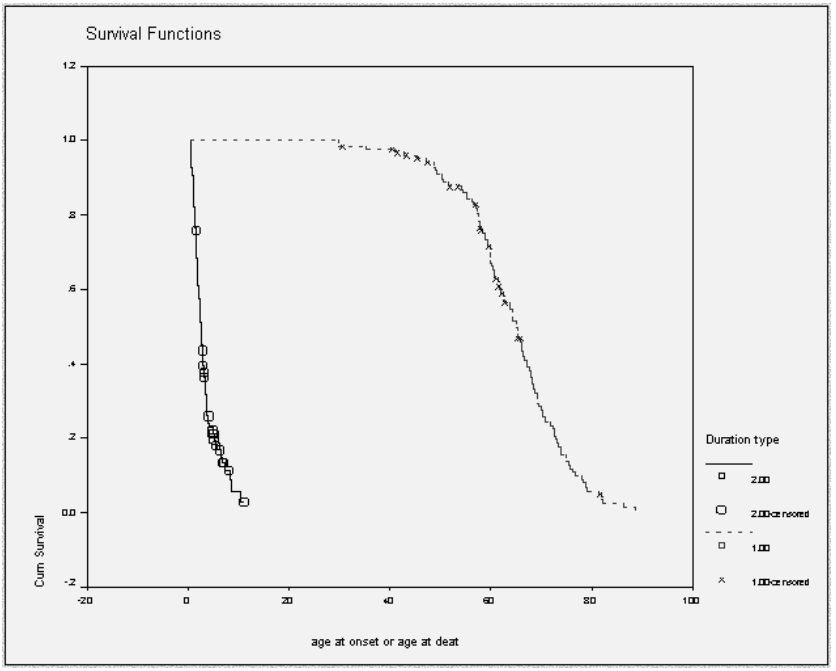


**Figure 1: Kaplan Meier estimate for age at onset and duration from onset to death, (time measured in years).**

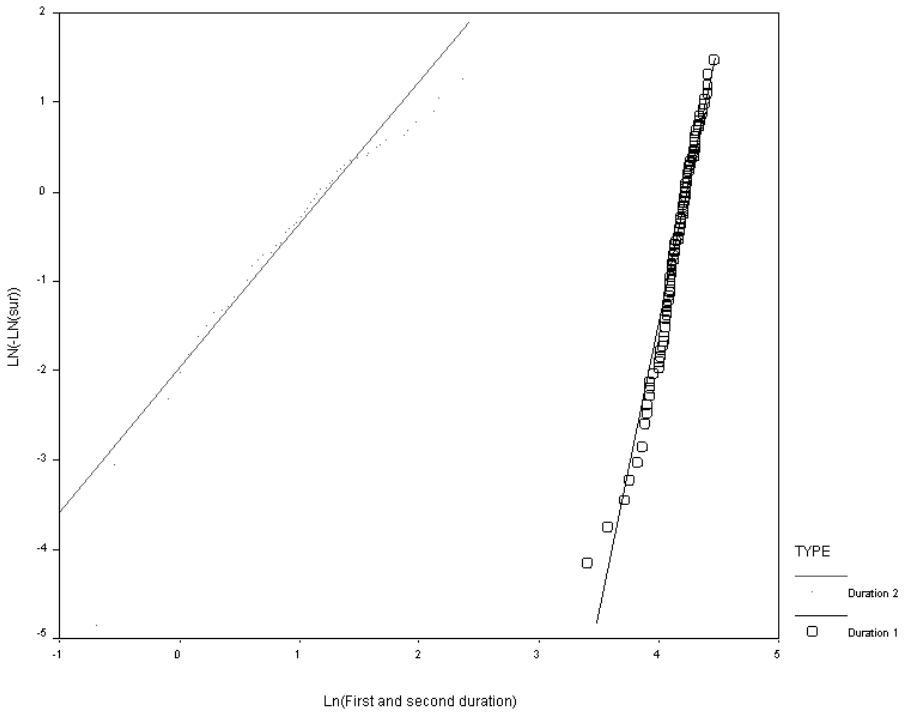


**Figure 2: complementary log-log transformation of Kaplan-Meier estimates for duration 1 (age at onset) and duration 2 (from onset to death), time measured in years.**

The main models were refitted for males and females separately; a gender effect in the prevalence of MND is one of the few universally agreed features of the disease. The means and standard deviations are shown in Table 6 calculated from the best fitting model in each case. The age of onset is estimated, on average, to be more than 2 years later for females. This is consistent with the theory that MND is a degenerative disease with the rate of degeneration varying over the population and many people not experiencing observable onset because they die of other causes. Under this theory, females would have a later average age of onset because they have a lower hazard of dying from other causes (longer life expectancy).

**Table 3: Modeling results for age at onset with no explanatory variables**

| Parameter | Weibull | | Log-logistic | | Lognormal | | Gompertz | |
|---|---|---|---|---|---|---|---|---|
| | Est | SE | Est | SE | Est | SE | Est | SE |
| Shape | 6.377 | .434 | 9.448 | .066 | 0.202 | .013 | 0.008 | .001 |
| Scale | -42.63 | 2.933 | 6.614 | 0.016 | 6.594 | .018 | -11.354 | .408 |
| LOG-LIKE | -808.05 | | -814.85 | | -820.65 | | -811.7 | |

**Table 4: Modeling results for age of death with no explanatory variables**

| Parameter | Weibull | | Log-logistic | | Lognormal | | Gompertz | |
|---|---|---|---|---|---|---|---|---|
| | Est | SE | Est | SE | Est | SE | Est | SE |
| Shape | 7.533 | 0.545 | 11.028 | 0.818 | 0.170 | .012 | 0.009 | .001 |
| Intercept | -50.880 | 3.707 | 6.691 | 0.144 | 6.681 | 0.016 | -12.509 | .510 |
| LOG-LIKE | -683.30 | | -686.5 | | -691.45 | | -686.6 | |

**Table 5: Modeling results for duration from onset to death with no explanatory variables.**

| Parameter | Weibull | | Log-logistic | | Lognormal | | Gompertz | |
|---|---|---|---|---|---|---|---|---|
| | Est | SE | Est | SE | Est | SE | Est | SE |
| Shape | 1.412 | .105 | 2.174 | .161 | 0.786 | .055 | 0.008 | .004 |
| Intercept | -5.368 | 0.427 | 3.438 | .071 | 3.446 | 0.071 | -3.996 | .149 |
| LOG-LIKE | -505.95 | | -498.92 | | -497.67 | | -564.10 | |

### *3.2.1 Methodology*

The relationships between the explanatory variables and durations were investigated using the parametric models in the previous subsection as a starting point (Fahrmeir et.al., 1994). This is a pragmatic approach designed to simplify model building into a sequence of relatively straightforward steps. It is recognized, however, that this approach carries a risk of not identifying the globally best fitting model. In particular, there is no guarantee that the parametric form of the survival distribution in the absence of covariates is also appropriate for the model with covariates. As we know, ignoring covariate effects will tend to result in a hazard which increases less steeply (or declines more steeply) than the "true" individual level hazard (see, for example, Lancaster T. et al., 1980).

Because of the large number of potential explanatory variables, we also adopt a pragmatic approach in selecting those to be included in the model. Specifically, we use the 'forward selection" successive model fitting procedure with the best additional variable added to the

model at each stage. The improvement in the model as a result of adding in each variable in turn was assessed by a likelihood ratio test and the process ceased when no additional variable was significant at the 5% level. In the first stage, we dropped non-significant variables and we included the most significant variable. Then in stage 2, we included the next most significant main variable and so on until no more variables were significant. Then we added interactions one at a time (although, in the event, none proved to be significant).

### *3.2.2 Results: Age of onset*

The results of the forward selection model fitting for age of onset are given in column 1 of Table 7. Occupational exposure to fumes and dust was the most significant variable and was therefore entered into the model at stage 1. The next variable to be included was glandular fever, with those who have had glandular fever having a higher hazard. However, the very low frequency of glandular fever in this dataset requires this result to be interpreted with caution. Working with industrial chemicals and solvents was the final variable to be entered but here is another problem: the significance is primarily due to a higher hazard for the "don't know" category. There are a number of possible explanations for this effect. With the small numbers involved, it might have occurred just by chance. If younger patients are experiencing more rapid deterioration, it may reflect the difficulties of obtaining complex information from patients in more advanced stages of the illness. On the other hand, most blue collar workers will have used industrial chemicals and solvents but, because this would usually be incidental to their work, they may have tended to reflect their uncertainty about the question by responding "don't know". If this theory is correct, the "don't know" category will be dominated by those with a low level of exposure. However, with the data available, we are unable to resolve this issue.

None of the interactions proved to be significant but two variables, significant at stage 1, ceased to be significant at later stages and therefore did not appear in the full model. These are exposure to laboratory chemicals and solvents and occupational exposure to rubber. That they reduced in significance when occupational exposure to fumes and dust and industrial chemicals and solvents were included

in the model, suggests that it is the general exposures rather than the specific industrial context which is important. This is consistent with the theory that the "don't know" group for industrial chemicals and solvents includes those with incidental exposure.

**Table 6: Means and Standard deviations from best fitting models (time measured in years).**

| | **Age of Onset** **Weibull** | | **Onset to Death** **Log-Normal** | | **Age of Death** **Weibull** | |
|---|---|---|---|---|---|---|
| | ***Mean*** | ***SD*** | ***Mean*** | ***SD*** | ***Mean*** | ***SD*** |
| Total | 64.17 | 10.75 | 3.55 | 2.99 | 67.08 | 10.53 |
| Male | 63.36 | 10.00 | 3.58 | 3.33 | 66.12 | 9.99 |
| Female | 65.64 | 14.78 | 3.47 | 3.08 | 68.42 | 10.95 |

**Table 7: Preferred models from forward substitution model fitting. Following convention, for the Weibull, the hazard is specified as a function of the explanatory variables while, for the log-normal, it is the mean which is specified as a function of the explanatory variables.**

| **Variable** | **Age of Onset** **Weibull** | | **Onset to Death** **Log-Normal** | | **Age of Death** **Weibull** | |
|---|---|---|---|---|---|---|
| | ***Estimate*** | ***SE*** | ***Estimate*** | ***SE*** | ***Estimate*** | ***SE*** |
| Shape | 7.053 | .486 | .715 | .050 | 8.398 | .608 |
| Scale | -45.02 | 3.218 | 5.230 | .384 | -54.17 | 0.577 |
| Age of Onset $\times 10^{-2}$ | | | -.185 | 0.05 | | |
| Indust. Chems. And Solvesnts: | | | | | | |
| No Exposure | -0.150 | 0.191 | | | -0.281 | .208 |
| Don't know | 1.505 | 0.612 | | | 1.539 | 0.619 |
| Lan. Chems. And Solvents: | | | | | | |
| No Exposure | | | -.473 | .198 | | |
| Don't know | | | -.380 | .401 | | |
| Fumes or Dust: | | | | | | |
| No Exposure | -0.582 | 0.199 | | | -0.613 | 0.220 |
| Don't know | -1.298 | 0.719 | | | -1.113 | 0.720 |
| Glandular Fever: | | | | | | |
| No Exposure | -1.846 | 0.604 | | | -2.189 | 0.61 |
| Don't know | -2.270 | 1.159 | | | -2.568 | 1.163 |
| LOG-LIKE | -796.45 | 30 | -485.79 | | -670.3 | |

### *3.2.3 Result: Age of Death*

The forward selection model fitting finished with the same set of variables included as for age of onset, and the variables enter in the

same order. The model fitting results are given in column 3 of Table 7. the similarities with age of onset are not surprising as age of death = age of onset + duration from onset to death and age of onset has more variability than the duration from onset to death.

***3.2.4. Results: Duration from onset to Death***

In interpreting these results, it is emphasized that the log-normal model is parameterized in a conventional formulation with the mean specified as a function of the linear predictor. For the Weibull, a positive parameter for a covariate indicates a positive relationship with the hazard and hence a negative relationship with the mean. For the log-normal, a positive parameter indicates a positive relationship with the mean.

The forward selection model fitting results for duration from onset to death are shown in column 2 of table 7. The first variable to be included (the most significant variable when included on its own in the model) is age of onset. Surprisingly, however, the parameter estimate for this variable is negative indicating a negative association between age of onset and the duration from onset to death; those with later onset tend to have shorter survival times than those with early onset. This is the opposite of what may be expected if the rate of deterioration theory discussed in section 1 is correct.

The only other variable which is significant in stage 1 is laboratory exposure to chemicals and solvents and this is added in to the model at stage 2. However, this result is also surprising because exposure is associated with longer and not shorter survival times. The implication is that chemicals and solvents are associated with earlier onset but slower progress of the disease after onset.

Hazards are plotted in figure 3 for three different ages of onset, with the reference category for laboratory exposure to chemicals and solvent. These plots show the substantially higher hazards for those with onset at age 60 compared to those with onset at age 40. One possible explanation for the observed relationship is that older patients tend to be more frail and therefore more vulnerable to the infections that can shorten the life of sufferers.

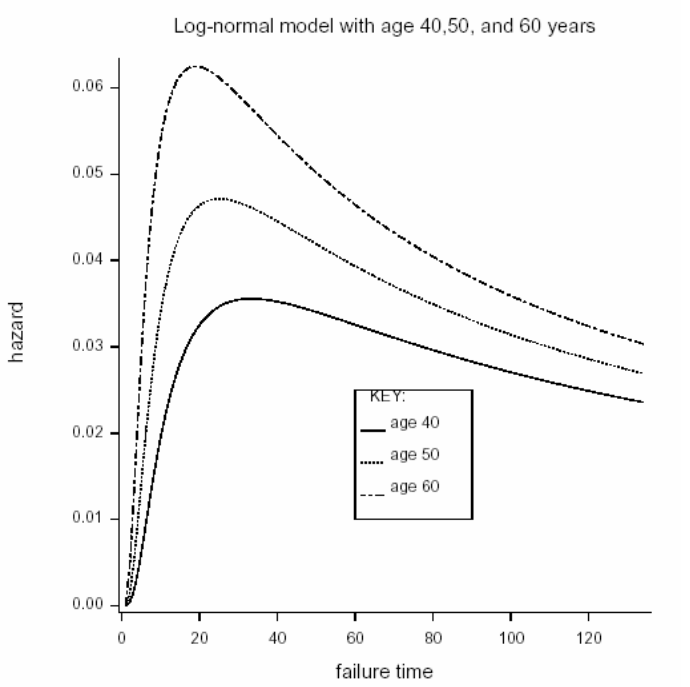


**Figure 3: Log-normal hazard for duration from onset to death for different ages at onset, (time measured in months).**

## 4. Two Stage Modeling

In this section we report the results of the joint modeling of age at onset and duration from onset to death. A conventional frailty formulation

$$f(t_1,t_2 \mid \mathbf{x},v) = f_1(t_1 \mid \mathbf{x},v) f_2(t_2 \mid \mathbf{x},v). \quad (1)$$

is not relevant because there appears to be a negative association between the two durations. We consider two approaches, one specific and one general, which enable frailty models to be extended to allow for negative associations. We consider also a Markov effect given by the decomposition

$$f(t_1,t_2 \mid \mathbf{x},v) = f_1(t_1 \mid \mathbf{x},v) f_2(t_2 \mid \mathbf{x},t_1,v). \quad (2)$$

### *4.1. Model Formulation*

First we consider a more general formulation for equation 2 given by

$$f(t_1,t_2 \mid \mathbf{x},\boldsymbol{\nu}) = f_1(t_1 \mid \mathbf{x},\nu_1) f_2(t_2 \mid \mathbf{x},t_1,\nu_2). \tag{3}$$

The integrated or marginal density now requires a bivariate integration

$$f(t_1,t_2 \mid \mathbf{x}) = \int\int f_1(t_1 \mid \mathbf{x},\nu_1) f_2(t_2 \mid \mathbf{x},t_1,\nu_2) dG(\nu_1,\nu_2). \tag{4}$$

This formulation may be motivated by arguing that, just as different combinations of covariates **x** may occur in the two linear predictors, so different combinations of the omitted variables may result in different, but possibly correlated, random effects in the two linear predictors.

#### *4.1.1 One-factor method of Heckman*

Following the method of (Heckman J. *et al.*, 1982), and writing $\nu_1 = c_1\nu$ and $\nu_2 = c_2\nu$, reduces equation 4 to a univariate integral corresponding to a conventional frailty formulation. With a hazard modeled for $T_1$ (Weibull) and a mean modeled for $T_2$ (log-normal), a positive association would imply that $c_1c_2 < 0$. By allowing $c_1c_2 > 0$, we extend the frailty approach to permit negative associations between durations.

For age at onset, $T_1$ is assumed to be Weibull with shape parameter $\gamma$ and location parameter $\lambda = \exp(\beta_0 + \boldsymbol{\beta}\mathbf{x}_{i1} + c_1\nu)$, $T_2$ is assumed to be log-normal with variance $\sigma^2$ and mean $\mu = \alpha_0 + \boldsymbol{\alpha}\mathbf{x}_{i2} + \psi t_1 + c_2$. Probability density functions for $T_1$ and $T_2$ and duration 2 are therefore given by

$$f_1(t_{i1} \mid \mathbf{x},\nu) = \lambda\gamma t_1^{\gamma-1} \exp[-\lambda t_1^{\gamma}],$$

and

$$f_2(t_{i2} \mid \mathbf{x},\nu) = \left[\frac{1}{\sqrt{2\pi}t_{i2}}\right] \exp\left\{-\frac{1}{2\sigma^2}(\log t_{i2} - \mu)^2\right\}.$$

Rewriting equation 4 and using quadrature to evaluate the integral gives the sample likelihood

$$L(\theta)=\sum_i\sum_k^m\left\{\begin{array}{l}\lambda\gamma t_{i1}^{\gamma-1}\exp(-\lambda t_{i1}^{\gamma})\\ \times\left[\dfrac{1}{\sqrt{2\pi t_2}}\right]\exp\left\{-\dfrac{1}{2\sigma^2}[\log t_{i2}-\mu]^2\right\}P_k\end{array}\right\}^{\delta_i}$$
$$\times\left\{\lambda\gamma t_{i1}^{\gamma-1}\exp(-\lambda t_{i1}^{\gamma})\left[1-\Phi\left(\frac{\log t_{i2}-\mu}{\sigma}\right)\right]P_k\right\}^{1-\delta_i}$$

where $\lambda=\exp(\beta_0+\boldsymbol{\beta}\mathbf{x}_{i1}+c_1\xi_k)$, $\mu=(\alpha_0+\boldsymbol{\alpha}\mathbf{x}_{i2}+\psi t_1+c_2\xi_k)$, $m$ is the number of quadrature points, $\Phi(.)$ is the distribution function of the standard normal, $\{\xi_k, P_k\}$ are the quadrature points, and $\delta_i$ is the censoring indicator for duration from onset to death, 1 for death and 0, otherwise.

We recognize that adopting these parameterization is a pragmatic, incremental approach which carries some risk of model misspecification; allowing for frailty effects could alter the nature of the duration distribution. There is therefore no assurance that the Weibull and log-Normal are the most appropriate parametric forms after including frailty effects. On the other hand, full frailty effect models are very demanding of computer time and are often difficult to fit because of problems in selecting appropriate starting values. It is therefore unrealistic to compare a number of different parameterization using a full frailty effects specification; some pragmatism is essential in using frailty models in realistic applications. For the same pragmatic reason, we use the covariates identified as significant in the conventional modeling.

### *4.2. Model Fitting Results*

Model fitting results are shown in Table 8 and 9 for six models in total. Introducing a frailty specification has only a modest effect on parameter estimates for covariates and these are omitted from the tables for brevity. Table 8 gives the results when the Markov dependence is excluded from the model. The first column gives the parameter estimates for the homogeneous formulation with no frailty

effects. This is included for completeness and corresponds to the two preferred models from the previous section.

The second column gives results for the single factor (scalar) model. As expected from the analysis in section 3, the scale parameter estimates for the two random effect terms have the same sign; omitted factors which increase the hazard and result in earlier onset tend also to increase survival times after onset. Consistent with theoretical expectations (see, for example, Lancaster T. *et al.,* 1980), the Weibull shape parameter increases with the introduction of the frailty effects. The changes to the log-normal hazard are not evident from the parameter estimates and, to aid the interpretation of the results, the two log-normal hazards are plotted in Figure 4. In plotting this Figure, the covariate is set to zero (the reference group) and, for the frailty model, the frailty effect is also set to zero. The theoretical expectation is that, with the inclusion of a frailty effect, the hazard will increase more steeply and decrease less steeply. Observation of Figure 4 suggests that this is true only over the central range of failure time; it is not the pattern at low and high failure times. This is presumably due to the set of hazard shapes permitted by a log-normal distribution being restricted. As expected, the conditional hazard peaks later than the marginal hazard.

**Table 8- Modeling two durations as Weibull-lognormal with and without random effect, not including Markov effect. NB: Parameter estimates for covariates are omitted for brevity.**

| Variable | | No | | Scalar | |
|---|---|---|---|---|---|
| | | Random Estimate | effect SE | Random Estimate | effect SE |
| Const | W | -45.034 | 3.218 | -50.714 | |
| Shape | W | 7.056 | 0.486 | 7.982 | |
| Const | LN | 4.015 | 0.160 | 3.916 | |
| SD | LN | .754 | 0.041 | .626 | |
| $c_1$ | | | | -.560 | |
| $c_2$ | | | | -.424 | |
| $q_1$ | | | | | |
| $q_2$ | | | | | |
| $q_3$ | | | | | |
| Log-Likelihood | | -1288.18 | | -1285.09 | |

**Table 9- Modeling two durations as Weibull-lognormal with and without random effect including Markov effect in all models. NB: Parameter estimates for covariates are omitted for brevity.**

| Variable | | No Random effect Estimate | No Random effect SE | Scalar Random effect Estimate | Scalar Random effect SE |
|---|---|---|---|---|---|
| Const | W | -45.034 | 3.218 | -46.893 | |
| Shape | W | 7.056 | 0.486 | 7.334 | |
| Const | LN | 5.229 | 0.384 | 5.756 | |
| SD | LN | .716 | 0.050 | .517 | |
| $10^{-2}\times$Age of Onset | | -.185 | .05 | -.225 | |
| $c_1u$ | | | | .247 | |
| $c_2u$ | | | | -.503 | |
| $q_1$ | | | | | |
| $q_2$ | | | | | |
| $q_3$ | | | | | |
| Log-Likelihood | | -1281.78 | | -1281.07 | |

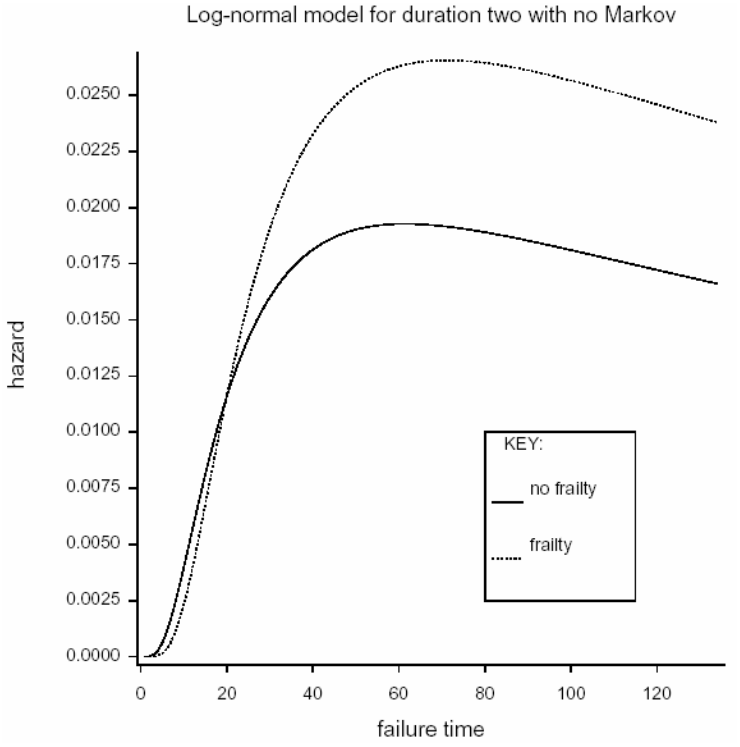

**Figure 4- Log-normal hazard for duration for duration from onset to death, with and without frailty, and with no Markov effect, time to measured in months**

Table 9 gives the results when the Markov dependence is included in the models. The parameter estimates for the homogeneous model with no frailty effects are shown in the first column. The improvement in the log-likelihood over the homogeneous model without the Markov effect is significant, with LR chi-square 12.80 ( $p > 0.001$ with 1 degree of freedom). As the likelihood function is separable, including the Markov effect does not change the parameter estimates for the Weibull model.

The parameter estimates for the scalar frailty effect model are shown in the second column of Table 9. An interesting feature of these results is that the Markov effect is intensified, counteracted by negatively correlated random effects; the two coefficients for the frailty terms have different signs. If this model is correct, this implies that there is a positive association between age of onset and duration from onset to death after allowing for the direct (i.e. Markov) effect of age of onset on the post-onset survival time. This positive association implies a frailty effect operating in the same direction for the two durations. We may therefore conceptualize the frailty effect as an underlying rate of deterioration with high frailty, for example, reducing age of onset and the duration from onset to death. Thus, in spite of the observed negative correlation between the two durations, it is possible that the underlying rate of deterioration in motor neuron function does mean that those with later onset will potentially live longer. However, this relationship is obscured by a separate age effect, older patients succumbing faster for reasons quite separate from underlying rate on motor neuron deterioration.

It is emphasized that this logic is speculative; it is impossible to draw firm conclusions because of the very similar effects of omitted variables and Markov-type dependence. This problem is demonstrated by the very small improvement in the log-likelihood, the LR chi-square is 1.42. ( $p = 0.49$, with 2 degrees of freedom, although a likelihood ratio test is not strictly appropriate because the homogenous model is at the boundary of the parameter space). The Weibull shape parameter increases as expected, although the increase

is more modest than in Table 8. This is presumably because of the lower frailty effect variances when the Markov effect is included in the model. The change in the log-normal hazard is shown in Figure 5. The age of onset is set at 50 years for these plots. The relative shapes of the conditional and marginal hazards are similar to the results for the models without a Markov effect. In particular, the conditional hazard increases more steeply and decreases less steeply than the marginal hazard throughout the central range of failure times. Also, the conditional hazard peaks later than the marginal hazard. However, the locations of both hazard maxima are notably later than for the models without a Markov effect.

## 5. Conclusions

The modeling results for univariate models suggest that:

i) There may be a connection between occupational exposure to fumes, dust, chemicals, and solvents and the age of onset of MND, with exposure related to earlier onset. Other studies have suggested a relationship between these types of occupational exposure and the occurrences of MND (Neilson S. Gunnarsson L. G Robinson I. 1994); we believe that this is the first study to relate such exposure to age of onset. We also found evidence that exposure to chemicals and solvents is associated with slower progression of the disease after onset.

ii)There may be a connection between glandular fever and earlier onset of MND, but we note the very low frequency of glandular fever in the dataset.

iii) Duration from onset to death is negatively related to age of onset. This finding challenges the theory that later onset is associated with a slow progression and early onset with a rapid progression of the disease. It also suggests that a conventional frailty model is inappropriate for modeling the dependence between these two durations.

The results for the simultaneous modeling of the two stages of the disease (onset and death) suggest that:

i) There is strong evidence of variation between individuals in the onset of MND and progression from onset to death, over and above that explained by covariates.

ii)Our results do not distinguish effectively between true and spurious duration dependence. However, there is a tantalizing suggestion from the full model with Markov dependence and frailty that there is a positive dependence consistent with the notion of MND arising from a progressive deterioration with a rate which varies over the population, and that the observed negative dependence is due to a different causal mechanism.

**Acknowledgments**

The authors are grateful to the Ministry of Culture and Higher Education of the Islamic Republic of Iran for their financial support. We are also grateful to Professor J.D. Mitchell for giving us the rich data about MND.

**Reference**

Buckley J. (1983) *Motor neuron disease in England and Wales*: 1950-1979/ J. Neurol. Neurosurg. psychiat., **46**, 197-205.

Lindley, D. and Singpurvalla N. (1986) *Multivariate distributions for the life length*. Journal of Appled Prob., **23**, 418-431.

Oakes D. (1989) *Bivariate survival models induced by frailties*. Journal of American Statistical Association, **84**, 487-493.

Ling Mei, and Lee Ting. (1993) *Mixtures of Lifetime Distributions*. Advances in Reliability Edited by Asit P. Basu. 231-245.

Hougaard P. (1995) *Frailty Models for Survival Data*. Lifetime data Analysis **1**, 255-273.

Xue S. and Brookmeyer R. (1996) *Bivariate Frailty Model For the Analysis of Multivariate Survival Time*. Lifetime Data Analysis. **2**, 277-289.

Yashin Anatoli I. and Iachine Ivan A. (1999) *What Difference Does the Dependence Between Durations Make?* Insights for Population Studies of Aging. Lifetime Data Analysis, **5**, 5-22.

Neilson S. (1993) *An explanation.* Acta Neurological Scandinavia: **87**, 184-191.

Neilson S. Robinson I. and Kondo K. (1993) *A New analysis of mortality from motor neuron disease in Japan, 1950-1990: rise and fall in the postwar years.* Journal of the Neurological Sciences, ***117,*** *46-53.*

Neilson S. Gunnarsson L. G. Robinson I. (1994) *Rising mortality form motor neurone disease in Sweden 1961-1990: the relative role of increase population life expectancy and environmental factors*. Acta Neurologica Scandinavica, **90**, 150-159.

Mitchell J. D., Davies R. B., Al-Hamad A. Gatrell A. C. and batterby G. (1995) *MND risk factors: an epidemiological study in the North West of England*. Journal of the neurological Sciences. **129**, 61-64.

Neilson S. Robinson I. and Hunter M. (1992) *Static and dynamic models of interdisease competition: Past and Projected Mortality from Amyotrophic Lateral Sclerosis and Multiple Sclerosis. Mechanisms* of Ageing and Development, Elsevier Scientific publisher Ireland ltd.

Neilson S. Robinson I. (1993) *Reinterpreting mortality statistics: Some uses of Gompertzian analysis in epidemiological research.* Journal of Epidemiology, **46**(9), 1063-1069.

Neilson S. Robinson, Robinson I. and Hunter M. (1993) *Static and dynamic Models of Interdisease Competition: Past and projected mortality from amyotrophic lateral sclerosis and Multiple sclerosis*. Mechanisms of Algeing and Development, **66,** 223-241.

Gatrell A.D., Sabel C.E., Nutcgekk H.D. (1999) *What's the Point? Exploring the Spatial Expidemiology of Motor Neurone Disease in Part of North-West England. Chapter 10 of Statistical Aspects of Health and Environment by Barnett. Stein and Turkman*. John Wiley and Sons ltd. Pages, 170-192.

Francis, B. Green, M. Payne, C. (1993) Glim4. *The Statistical System for Generalized Linear Interactive Modelling*. Clarendon Press. Oxford.

Lancaster T. and Nickel S. (1980) *The analysis of Te-Employment Probabilities for the Unemployed.* Journal of the Royal Statistical Society. A **134**, 141-165.

Hechman, J.J., and Singer. B. (1982) Population heterogeneity in demographic models. *Multidimensional Mathematical Demography,* 567-595.

Fahrmeir L. and Tutz G. (1994) *Multivariate Statistical Modelling Based on Generalized Linear Models. With Contributions by Wolfgang Hennevogl*. Springer Verlag.

Dos Santos, D. (1994) *Distinguishing the Effects of time Dependence from the Effects of Frailty in Multiple Spell Breast Cancer data. Thesis,* Cenctre for Applied Statistics, Lancaster University.