8-2011

# Stressors Across the Lifespan and Dementia Risk: A Statistical Method Analysis

Megan Platt Borrowman
*Utah State University*

Stressors Across the Lifespan and Dementia Risk:
A statistical method analysis

By

Megan Borrowman

A report submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Statistics

Approved:

______________________________ ______________________________

| | |
|---|---|
| Dr. Christopher Corcoran | Dr. Maria Norton |
| Major Professor | Committee Member |

______________________________

Dr. Adele Cutler
Committee Member

UTAH STATE UNIVERSITY
Logan, Utah

2011

# Stressors Across the Lifespan and Dementia Risk:

## A Statistical Method Analysis

Megan Borrowman
Plan B Thesis Project

Major Professor: Christopher Corcoran
Department: Mathematics and Statistics

## Abstract

The Cache Lifespan Stressors and Alzheimer's Disease (LSAD) study has access to data from the Cache County Study on Memory Health and Aging (CCS) that have been linked to the extensive genealogical and vital records from the Utah Population Database (UPDB). Information about stressful life events experienced by the original 5092 CCS participants has been extracted objectively from the UPDB, without the possibility of recall bias. This information was then statistically analyzed to look for relationships between key stressors and dementia risk. The LSAD study made it possible to examine the correlation between stressors as well as look at patterns or groupings that may exist among the stressors.

For this project, we will apply, compare and contrast a variety of methods to explore the relationship between stressors and dementia status. We first assessed significant associations between stressors and dementia using chi-square tests, logistic regression, and stepwise logistic regression. CART and Random Forests were also used to evaluate which variables are most predictive of dementia status. To explore interrelationships among the stressors, principal components analysis, factor analysis, and cluster analysis were performed. These last methods also obtain dimension reduction of the set of stressor variables to a smaller set of factors or clusters of stressors. For each of the methods, we explored some basic strategies to account for the high rate of missingness among some of the stressors.

8/10/11 4:46 PM

**Comment:** Is it appropriate to use the first person plural since this is the student's work, primarily? This makes it sound as though it was a research team with multiple co-authors/co-investigators. Is active or passive voice appropriate? For example, could change this sentence to: "A variety of methods were applied and compared, to explore the relationship……"

8/10/11 4:46 PM

**Comment:** Probably need to make everything paste tense since this is a report, after work was completed.

Logancity 8/10/11 5:06 PM

**Comment:** I thinks it is appropriate to use "we" in this paper. The only time I used future tense is in the introduction.

## Introduction

The proportion of elderly adults is expected to increase significantly both in the U.S. and worldwide.[1] With that increase the number of adults with dementia will also grow. The costs of treatment and caregiving for individuals with dementia are significant, and the growing prevalence of dementia represents an increasingly enormous health burden. There is a pressing need to explore and better understand the causes and risk factors associated with dementia, to provide better means of prevention, diagnosis, and treatment.[3]

Prior research has long suggested that stressful life events may have a significant impact on outcomes in late life. This relationship led to the development of the so-called Holmes-Rahe Stress Index, which has been used to evaluate a person's risk of various illnesses such as headache, diabetes, fatigue, hypertension, chest and back pain, ulcers, infectious disease based on the stressful events, or stressors, they have experienced in the last 1-2 years.[2] Though this measure is widely used throughout the social sciences, the short amount of time it covers is a considerable limitation.

With respect specifically to dementia, research has shown that high levels of glucocorticoids released in the brain during stressful situations can lead to damage of parts of the brain responsible for learning and memory.[5] The neural degeneration resulting from the damage to the brain due to the response to stress may increase susceptibility to dementia. This etiology is consistent with findings from various observational studies, which have demonstrated an association between dementia risk and stressors such as death of a parent during childhood,[6] death of a spouse, and death of a child.[3]

The Cache County Lifespan Stressors and Alzheimer's Disease (LSAD) study provides an important opportunity to assess the relationship between early life stressors and eventual dementia risk. This study combines data from two rich resources: the Cache County Study on Memory Health and Aging (CCS), along with linkage of the Cache cohort to the Utah Population Database (UPDB). The comprehensive genealogical and vital records of the UPDB provide a unique resource that eliminates the subjectivity of a personal survey regarding early life events, ensuring no recall bias. The CCS is a longitudinal population-based study initiated with NIH funding in 1995, involving 5092 original Cache County, UT, residents aged 65 years or older at study entry. At each of three triennial follow-ups after baseline assessment, surviving CCS participants without a dementia diagnosis in a prior assessment have been evaluated with respect to cognitive decline and clinically verified dementia onset, along with a host of other demographic, health, genetic, and other risk factor variables.[4] As part of an NIH-funded project in 2000, ninety-nine percent of the records from the CCS were successfully linked with records in the UPDB.[3] The UPDB contains extensive vital records, including birth certificates, death certificates and marriage licenses. This provides a unique opportunity to objectively evaluate stressors *across the lifespan* and their association with dementia risk.

Although the several stressors available through LSAD can be examined individually, the objective of this project is to examine them collectively, both relative to their associations with dementia and to each other. We are therefore interested in multivariate approaches that will identify novel patterns among the stressors both for the purpose of data reduction and predictive power. These methods will be informed by univariate and multivariate exploratory analyses, and we will compare and contrast the prediction and data reduction results with conventional modeling strategies such as logistic regression. In addition, the problem of missing data is common to all of these analyses, due to the relatively high rate of missing values among some of the stressors. Various approaches of handling the missing value problem will be applied to each statistical method and the results will be compared.

## Methods

The purpose of this project is to understand the relationship between dementia and stressful events across the lifespan. In this study, both prevalent and incident dementia were classified as demented. In univariate analyses of the individual stressors, it was determined that only 9 of the 14 stressors had sufficiently high prevalence to have any meaningful statistical power. Thus, the methods described below were applied only to the 9 most prevalent stressors, which include: death of mother during childhood, death of father during childhood, offspring death, sibling death, low birthweight or premature offspring, divorce, widowhood, low socioeconomic status (SES), and low education.

In examining associations between these stressors, our initial bivariate analyses used standard methods for rates and proportions, including chi-square and Fisher's exact tests (for those stressors with relatively smaller samples sizes)

8/10/11 5:01 PM
**Comment:** It probably needs to be stated in this section whether or not you excluded prevalent dementia cases, or included all dementia cases, along with rationale for that decision, as this has implications for interpretation of your findings.

Logancity 8/10/11 5:02 PM
**Comment:** Comment 3 addressed here.Thoughts on rationale?

to assess whether risk of dementia differed significantly between those with a given stressor and those without.

These analyses were followed by modeling and classification approaches – described briefly below – both to identify stressors with significant effects in the presence of the others (using logistic regression), and stressors identified as most important in predicting dementia status (using classification and regression trees along with random forests). We additionally considered the role of missing stressor data across these approaches.

In addition, we applied multivariate techniques in order to explore interrelationships between the stressors. We used principal components analysis to compare the similarities and differences between stressors and reduce them to a smaller number of components related to the original stressors. We also used factor analysis to identify hidden or unobserved traits that account for the correlation between stressors. Finally, we also used cluster analysis to split the data into correlated groups to examine the structure of the data.

*Logistic regression*

Logistic regression (LR) is a generalized linear model for binomial data that uses a logit link. We first applied LR to the 9 stressors individually. Fitted LR models yield a coefficient and likelihood-based p-value associated with the stressor of interest. A convenient aspect of these coefficients is their interpretability on the log-odds scale – positive coefficients indicate increased risk and negative coefficients indicate decreased risk.

A common programming default for many statistical procedures in the analysis environment R[7] is that the procedure fails if missing values are included in the investigative data set. As every stressor was missing for at least some subjects (e.g., nearly 30% of the data involving the offspring of the subject was missing), we applied three different approaches to address this. First, Method 1 involved dropping entire observations or subjects with any missing values. Our second approach was to first entirely drop stressors with high rates of missingness, such as offspring death, before eliminating those subjects with missing values for the remaining stressors. Our third strategy was to simply treat "missing" for a given stressor as an additional category, treating each stressor as a multinomial three-level variable (i.e., present, absent, or missing).

*Stepwise logistic regression*

Stepwise logistic regression is an automated model selection method that sequentially adds or subtracts variables to or from the regression model until a model remains containing only coefficients that have significant associations with the outcome in the presence of the others. The result of stepwise logistic regression is the predictor regressed upon the best combination of variables to achieve the most negative Akaike Information Criterion (AIC).

For the LSAD data, the three types of stepwise logistic regression (forward, backward, and combined forward and backward) were applied, using the three missing data strategies described earlier. The final model, the one with the most negative AIC, from each regression type was extracted and the variables included were evaluated.

The model achieved through stepwise logistic regression, often with variable reduction capabilities, is desirable, but it is important to consider the shortcomings of the method. Though it can successfully handle structural data problems such as collinearity, the instability of the method is a significant weakness. Even a small change in the data usually changes the model which, in turn, affects the interpretation. There can be additional difficulty in interpreting the p-values, which lose their meaning after stepwise regression has been performed.

CA*RT*

As a means of identifying predictors with the most predictive power, we first applied classification and regression tree (CART), a tree-based method. Although CART provides excellent interpretability and visual results, it is relatively less stable, and better accuracy can often be achieved by more sophisticated tree-based methods. CART considers all possible splits for the data and makes a selection using Gini criterion. This process of splitting, or branching, continues until homogenous ending nodes are achieved. The visual results show a classification tree which can be readily interpreted by following the path of the node.

*Random Forests*

Random Forests is another tree-based method that grows many regression or classification trees which are then tested using a bootstrap sample. A new object is classified based on the "votes" of the classification trees. The measure of the fit of the model is given by the out-of-bag (oob) error rate. Once the data has been passed down the trees, proximities are calculated that can be used to create visualizations that aid in interpretation. Random Forests also result in variable importance plots, yielding added interpretability through visualization. Random Forests have proven to improve accuracy relative to CART alone, and have the additional property that an arbitrarily large number of trees does not result in overfitting.

*Factor Analysis*

In general, factor analysis is used to uncover hidden or latent factors that explain the correlation between variables, or in our case, stressors. Factor analysis attempts to account for correlation structure among variables by dividing them up so that variables in the same group, or factor, are highly correlated and the variables in different groups are not highly correlated.

*PCA*

A statistical method applied to the LSAD data to explore the relationship between stressors is principal components analysis (PCA). The goal of PCA is dimension reduction. An advantage of PCA is it can handle a large amount of variables. It takes those variables and reduces them down to a small number of principal components. The first principal component has the most variation; the second principal component has the second most variation, and so on. The result is a linear combination of all of the original variables. A data assumption PCA has is that variability is information. Also, if the variables are measured on different scales, the data needs to be standardized with mean zero and variance of one before the analysis can be performed. There is the risk of losing information when combining so many variables together to form principal components. Another consideration when using PCA is that it is very affected by outliers.

The statistical methods of stepwise logistic regression, CART, Random Forests, and principal component analysis all have arguments to handle missing data. For all four methods, the default is for the method to fail if there are missing values involved. Once the argument was changed to exclude the subjects who had missing values for any of the stressors, the method succeeded and results were obtained.

*Cluster analysis*

Another method used to explore the interrelationship between the stressors is cluster analysis. The goal of cluster analysis is to split the data into groups or clusters where the observations within each group are similar to each other and different from the observations in other groups. In cluster analysis, the groups are determined based on similarities between the observations, whereas factor analysis, as described previously, uses similarities among the variables to determine groupings. There are two types of cluster analysis, hierarchical and partitioning. We will be using hierarchical methods that split the data into clusters by a series of steps. Agglomerative hierarchical clustering is the most common type and is the process of grouping observations together that are the "closest." The way the distance, or closeness, is calculated is different depending on whether single, complete, or average linkage is used. Another option is Ward's method that joins the clusters together in such a way as to result in the smallest within-cluster sum of squares error. A disadvantage of cluster analysis is the availability of multiple clustering methods, but no clear way to select the most appropriate one. Another weakness of clustering methods is that the number of clusters the data is split into is critical for maximizing interpretability, but it is difficult to determine from the data.

## Results

### Descriptive variables, stressors, and dementia

As noted earlier, the LSAD data extracted 14 stressful life events from the UPDB. T-tests and chi-square tests were used as an exploratory tool in determining the marginal associations between dementia risk, demographic factors, and stressors. Tables 1 and 2 contain descriptive statistics respectively for demographic factors such as age, education, gender, and APOE genotype, along with the investigative stressors.

Table 1: Descriptive statistics for continuous variables

| **Factor** | | **Range** | **Mean(s.d)** | **Median** | **p-value** |
|---|---|---|---|---|---|
| Age at baseline | | | | | |
| | Demented | 65-105 | 79.3(7.3) | 80.0 | <0.001 |
| | Non-demented | 65-103 | 75.0(7.1) | 74.0 | |
| | | | | | |
| Years of Education | | | | | |
| | Demented | 0-20 | 12.9(3.0) | 12.0 | 0.011 |
| | Non-demented | 0-20 | 13.2(2.7) | 12.0 | |
| | | | | | |

Both age and education were found to have statistically significant differences in subjects with and without dementia. Those subjects who were older at baseline were more likely to have dementia than younger subjects. The significant p-value of 0.011 for education indicates that those with less education were more at risk for dementia than those subjects with more education. Age may be influencing the association education has with dementia risk. In future studies, it would be important to determine whether the subjects who were older at baseline tended to have lower education.

Table 2: Binary variables and stressors with dementia rates from chi-square tests

| **Factor** | | **n(%)** | **% Demented** | **p-value** |
|---|---|---|---|---|
| Status | | | | |
| | Any Dementia | 942(18.5) | | |
| | Non-Demented or Missing | 4150(81.5) | | |
| | | | | |
| Gender | | | | |
| | Male | 2164(42.5) | 16.7 | 0.006 |
| | Female | 2928(57.5) | 19.8 | |
| | Missing | 0(0.0) | | |
| e4 allele | | | | |
| | With | 1604(65.9) | 27.7 | >0.001 |
| | Without | 3358(31.5) | 14.5 | |
| | Missing | 130(2.6) | | |
| Death of mother during childhood (ChildMaDeath) | | | | |
| | Yes | 335(6.6) | 24.2 | 0.007 |
| | No | 4749(93.3) | 18.1 | |
| | Missing | 8(0.2) | | |
| Death of father during childhood (ChildPaDeath) | | | | |
| | Yes | 426(8.4) | 22.3 | 0.041 |
| | No | 4652(91.4) | 18.2 | |
| | Missing | 14(0.3) | | |
| Sibling death (SibDeath) | | | | |
| | Yes | 3887(76.3) | 18.3 | 0.627 |
| | No | 1204(23.6) | 19.0 | |
| | Missing | 1(0.0) | | |

Table 2 continued: Binary variables and stressors with dementia rates from chi-square tests

| Factor | | n(%) | % Demented | p-value |
|---|---|---|---|---|
| Offspring death (OffDeath) | | | | |
| | Yes | 529(10.4) | 22.9 | 0.014 |
| | No | 3150(61.9) | 18.3 | |
| | Missing | 1413(27.7) | | |
| Low birthweight or premature offspring (lbwp) | | | | |
| | Yes | 478(9.4) | 15.5 | 0.084 |
| | No | 4613(90.6) | 18.8 | |
| | Missing | 1(0.02) | | |
| Adolescent parenthood (TeenParent) | | | | |
| | Yes | 33(0.6) | 21.2 | 0.906 |
| | No | 3656(71.8) | 18.9 | |
| | Missing | 1403(27.6) | | |
| | | | | |
| Old age parenthood (OldParent) | | | | |
| | Yes | 98(1.9) | 16.3 | 0.598 |
| | No | 3591(70.5) | 19.0 | |
| | Missing | 1403(27.6) | | |
| Divorce (divorce) | Yes | 157(3.1) | 10.8 | 0.012 |
| | No | 4566(89.7) | 19.1 | |
| | Missing | 369(7.2) | | |
| Widowhood (Widow) | | | | |
| | Yes | 2441(47.9) | 17.9 | 0.087 |
| | No | 2282(44.8) | 19.9 | |
| | Missing | 369(7.2) | | |
| Non-married at birth of first child (nomaroff) | | | | |
| | Yes | 18(0.4) | 27.8 | 0.505 |
| | No | 4705(92.4) | 18.8 | |
| | Missing | 369(7.2) | | |
| Never married (nomar) | | | | |
| | Yes | 81(1.6) | 13.9 | 0.279 |
| | No | 4642(91.2) | 19.0 | |
| | Missing | 369(7.2) | | |
| Married but no offspring (marnooff) | | | | |
| | Yes | 82(1.6) | 25.6 | 0.152 |
| | No | 4641(91.1) | 18.7 | |
| | Missing | 369(7.2) | | |
| Low SES (lowses) | | | | |
| | Yes | 1074(21.1) | 19.1 | 0.385 |
| | No | 3797(74.6) | 18 | |
| | Missing | 221(4.3) | | |
| Low education (lowed) | | | | |
| | Yes | 943(18.5) | 21.8 | 0.004 |
| | No | 4131(81.1) | 17.7 | |
| | Missing | 18(0.4) | | |

The results of the chi-square tests show that the variables gender and e4 allele are highly statistically significant descriptive variables. Due to insufficient sample sizes, the following stressors were not used in our analyses: adolescent parenthood, old age parenthood, non-married at birth of first child, never married, and married but no offspring. Of the remaining 9 stressors, death of mother during childhood, death of father during childhood, offspring death, divorce, and low education were significantly associated with dementia risk.

Logistic regression was next individually and collectively applied to the 9 most prevalent stressors. Results from the individual model fits and the fit with all stressors simultaneously are contained in Table 3.

Table 3: Logistic regression results

| **Stressor** | **p-value 1** | **Odds Ratio 1*** | **p-value 2** | **Odds Ratio 2**** |
|---|---|---|---|---|
| Death of mother during childhood | 0.006 | 1.44 | 0.255 | 1.20 |
| Death of father during childhood | 0.035 | 1.29 | 0.097 | 1.28 |
| Sibling death | 0.6 | 0.96 | 0.898 | 1.01 |
| Offspring death | 0.012 | 1.33 | 0.007 | 1.38 |
| Low birthweight or premature offspring | 0.074 | 0.79 | 0.053 | 0.74 |
| Divorce | 0.010 | 0.51 | 0.025 | 0.52 |
| Widowhood | 0.080 | 0.88 | 0.159 | 0.88 |
| Low SES | 0.361 | 1.08 | 0.511 | 1.07 |
| Low education | 0.004 | 1.30 | 0.033 | 1.27 |

*Odds ratio for separate logistic regression models for each stressor
**Odds ratio for logistic regression for the full model with all 9 stressors included

The stressors that are highly statistically significant are the following: death of mother during childhood, offspring death, divorce, and low education. Death of father during childhood was also significant. Just as in the chi-square tests, both low birthweight or premature offspring and widowhood were marginally non-significant.

The results from the logistic regression models indicate the influence each of the stressors has on dementia risk. Similar results were achieved from the individual and the 9-stressor model with the exception of death of mother during childhood. The individual model suggests those who experienced of a mother during childhood have 1.44 times the odds of dementia compared to those without death of a mother whereas the odds decrease to 1.20 in the full 9-stressor model. Death of a father during childhood, offspring death, and low education all have about 1.3 times the odds of having dementia compared to those without the stressor. Divorce was the only significant stressor with a protective effect. Subjects who were divorced had only about 50% the risk of dementia compared to those who did not experience the stressor of divorce.

Stepwise logistic regression has three different automated approaches to creating the final model. Forward stepwise logistic regression adds stressors to the model until the one most predictive of dementia status is achieved. Backward stepwise logistic regression begins with the full model and drops stressors until the final model is reached. There is also the option of a combination of forward and backward stepwise logistic regression that adds and/or subtracts stressors at each step to arrive at the final model with the lowest AIC. The three strepwise logistic regression approaches were applied to the three missing methods, resulting in a total of nine stepwise logistic regression models.

Table 4: Stepwise logistic regression models for Missing Method 2

| Method | Stressors in model | Odds Ratio | 95% C.I. | p-value |
|---|---|---|---|---|
| Forward | Death of mother during childhood | 1.37 | (1.03, 1.80) | 0.027 |
| | Death of father during childhood | 1.31 | (1.01, 1.68) | 0.038 |
| | Sibling death | 0.95 | (0.79, 1.15) | 0.613 |
| | Low birthweight or premature offspring | 0.80 | (0.61, 1.04) | 0.108 |
| | Divorce | 0.53 | (0.31, 0.86) | 0.015 |
| | Widowhood | 0.83 | (0.72, 0.98) | 0.022 |
| | Low SES | 1.04 | (0.86, 1.24) | 0.694 |
| | Low education | 1.27 | (1.05, 1.54) | 0.014 |
| | | | | |
| Backward and Combination | Death of mother during childhood | 1.32 | (1.00, 1.73) | 0.044 |
| | Death of father during childhood | 1.32 | (1.03, 1.68) | 0.025 |
| | Sibling Death | 0.93 | (0.77, 1.11) | 0.430 |
| | Low birthweight or premature offspring | 0.79 | (0.60, 1.02) | 0.076 |
| | Divorce | 0.51 | (0.30, 0.83) | 0.010 |
| | Widowhood | 0.86 | (0.74, 1.00) | 0.049 |
| | Low education | 1.29 | (1.07, 1.54) | 0.006 |

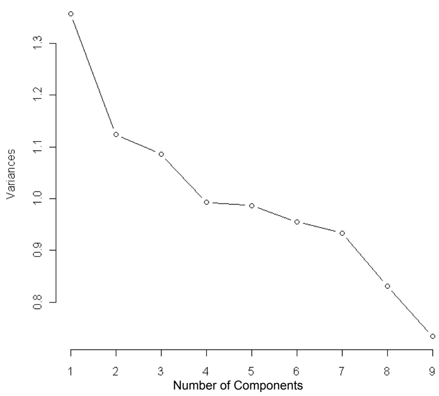
The stressors that were present in all nine models included death of father during childhood, low birthweight or premature offspring, divorce, and low education. The stressors that appeared in nearly all of the models were death of mother during childhood and widowhood.

Given potential computational and interpretability issues with generalized linear models, an alternative way to assess which stressors are most predictive of dementia status is the method of classification and regression trees (CART). CART results in a visual classification or regression tree that identifies which variables are responsible for the splits in the data. These variables help identify whether or not a subject has dementia. Unfortunately, due to the nature of the LSAD data, CART could not identify good splits to make a classification tree unless the complexity factor was set to an unreasonably low value. The associations the stressors were previously found to have with dementia were not strong enough to result in a classification tree. Since Random Forests is a collection of classification trees, naturally we were not able to obtain satisfactory results due to the low predictive power of the stressors. The associations found using chi-square tests and logistic regression were too weak to use for prediction.

**Between-stressor associations**

Principal components analysis is the first method applied to the LSAD data to assess the associations that may exist between each of the stressors. PCA considers the variances of each of the variables and typically reduces them to a smaller number of components. We first generated a scree plot to determine the number of components, contained in Figure 1. The shape of the curve suggests two components.

Figure 1: Scree plot for principal components




The two principal components were associated with different stressors depending on which method handling missing values was used. The principal components analysis results for each of the three missing methods are contained in Tables 5-7.

Table 5: Principal components analysis results for Missing Method 1

| Stressor from MM1 | PC1 | PC2 |
|---|---|---|
| Death of mother during childhood | -0.215 | 0.122 |
| Death of father during childhood | -0.187 | -0.082 |
| Sibling death | -0.309 | -0.376 |
| Offspring death | -0.413 | -0.272 |
| Low birthweight or premature offspring | 0.033 | -0.119 |
| Divorce | -0.002 | 0.121 |
| Widowhood | -0.386 | -0.562 |
| Low SES | -0.429 | 0.543 |
| Low education | -0.564 | 0.351 |

For Missing Method 1, the first principal component (PC1) is negatively associated with all stressors except low birthweight or premature offspring which has a very weak positive association. The strongest negative associations are with socioeconomic (SES) variables as well as sibling death and widowhood. This suggests that PC1 is a measure of high SES. The second principal component (PC2) has the strongest associations with widowhood and low SES being negative and positive respectively. Though weaker, PC2 is also associated negatively with offspring death and positively with low education. PC2 may be a measure of having your spouse but not being financially well off.

Table 6: Principal components analysis results for Missing Method 2

| Stressor from MM2 | PC1 | PC2 |
|---|---|---|
| Death of mother during childhood | -0.224 | 0.053 |
| Death of father during childhood | -0.173 | -0.093 |
| Sibling death | -0.410 | -0.456 |
| Low birthweight or premature offspring | 0.007 | -0.023 |
| Divorce | -0.032 | 0.358 |
| Widowhood | -0.314 | -0.643 |
| Low SES | -0.537 | 0.404 |
| Low education | -0.602 | 0.273 |

Table 7: Principal components analysis results for Missing Method 3

| Stressor from MM3 | PC1 | PC2 |
|---|---|---|
| Death of mother during childhood | -0.011 | 0.265 |
| Death of father during childhood | -0.008 | 0.248 |
| Sibling death | -0.332 | 0.255 |
| Offspring death | -0.400 | -0.114 |
| Low birthweight or premature offspring | -0.139 | 0.009 |
| Divorce | 0.656 | 0.036 |
| Widowhood | 0.527 | 0.230 |
| Low SES | -0.036 | 0.612 |
| Low education | 0.023 | 0.601 |

The results of the other two missing methods had more clear interpretations of the principal components. For Missing Method 2, the first principal component has a strong negative association with both SES stressors and a slightly weaker negative association with sibling death. This indicates PC1 is a measure of high SES. The second principal component has a strong negative association with widowhood and sibling death. PC2 is also positively associated with divorce and low socioeconomic status. This principal component could be an indicator of lack of close family deaths. Missing Method 3 resulted in the first principal component having a strong positive association with both widowhood and divorce. This PC is a measure of loss of a spouse. The second principal component has a strong positive association with both socioeconomic stressors. This indicates PC2 is measuring low SES.

Factor analysis was applied to the LSAD data and it was found that three factors accurately capture the correlation groupings. The factor loadings resulting from the factor analyses are contained in Tables 8 and 9.

Table 8: Factor loadings for Missing Method 1

| Stressor from MM1 | Factor 1 | Factor 2 | Factor 3 |
|---|---|---|---|
| Death of mother during childhood | | | 0.104 |
| Death of father during childhood | | | |
| Sibling death | 0.138 | | 0.126 |
| Offspring death | 0.112 | | |
| Low birthweight or premature offspring | | 0.900 | |
| Divorce | | | |
| Widowhood | 0.910 | | |
| Low SES | | | 0.385 |
| Low education | | | 0.593 |

Table 9: Factor loadings for Missing Method 2

| Stressor from MM2 | Factor 1 | Factor 2 | Factor 3 |
|---|---|---|---|
| Death of mother during childhood | 0.106 | | |
| Death of father during childhood | | | |
| Sibling death | 0.112 | 0.166 | |
| Low birthweight or premature offspring | | | 0.422 |
| Divorce | | | |
| Widowhood | | 0.666 | |
| Low SES | 0.382 | | |
| Low education | 0.584 | | |

The factor loadings indicate which factor each stressor is correlated with and the strength of the correlation. Using Missing Method 1, factor 1 is very highly correlated with widowhood (r = 0.910), factor 2 is very highly correlated with low birthweight or premature offspring, and factor 3 is correlated with both SES stressors. The distinct factors achieved is an ideal result from factor analysis. Factor 1 for Missing Method 2 is correlated with both SES stressors, factor 2 is correlated with widowhood, and factor 3 is correlated with low birthweight or premature offspring. Missing Method 3 would require too many variables to get any groupings or dimension reduction from factor analysis.

Factor loadings from Missing Method 2 were extracted and logistic regression was then applied to assess the association between the factor loadings and dementia status. This additional analysis is an extension of the between-stressor analysis done using factor analysis. Table 10 contains the results of the logistic regression analysis of the factor loadings from Missing Method 2.

Table 10: Logistic regression results of the Missing Method 2 factor loadings

| Factor | Odds Ratio | p-value |
|---|---|---|
| 1 | 1.16 | 0.009 |
| 2 | 0.90 | 0.060 |
| 3 | 0.81 | 0.029 |

Factor 1 of Missing Method 2 was statistically significant indicating those subjects for whom Factor 1 is present have about 1.16 the odds of dementia compared to those who do not have the factor. Factor 3 was also statistically significant, though subjects with Factor 3 have decreased odds, about 81% the risk of dementia compared to those who do not have the factor. Factor 2 was marginally non-significant.

Ward’s method was used to perform cluster analysis to look for patterns or groupings in the data. This approach appeared to yield no distinct clusters that would yield insights into the aggregation of study participants into distinct subgroups.

## Discussion

These analyses suggest that the death of mother during childhood, death of father during childhood, offspring death, low birthweight or premature offspring, divorce, and low education are all associated with dementia risk, although the associations appear insufficiently strong to use for prediction. The data reduction techniques most often identify some relationship between the two SES stressors. Stressors involving spousal events also appeared to be correlated. In all of the analyses, sibling death seemed to have the smallest impact on dementia status or stressor correlation of any of the stressors. Additional studies on the LSAD data should compare our results to those with a three-category outcome (no dementia, incident dementia, and prevalent dementia.) Future analyses should further focus on the modifying effects of age, gender, APOE status, education, and other key risk factors.

## References


1. Goulding, M., M. Rodgers, and S. Smith, *Trends in aging--United States and worldwide.* MMWR, 2003. 52(6): p. 101-4, 106

2. Holmes, T.H. and R.H. Rahe, *The Social Readjustment Rating Scale.* Journal of psychosomatic research, 1967. 11(2): p. 213-8

3. Lifespan Stressors and Alzheimer's Disease, grant number NIH AG031272

4. Breitner, J.C., B.W. Wyse, J.C. Anthony, K.A. Welsh-Bohmer, D.C. Steffens, M.C. Norton, J.T. Tschanz, B.L. Plassman, M.R. Meyer, I. Skoog, and A. Khachaturian, *APOE-epsilon4 count predicts age when prevalence of AD increases, then declines: the Cache County Study.* Neurology, 1999. 53(2): p. 321-31

5. Sapolsky, R.M., L.C. Krey, and B.S. McEwen, *The neuroendocrinology of stress and aging: the glucocorticoid cascade hypothesis.* Endocr Rev, 1986. 7(3): p. 284-301

6. Persson, G. and I. Skoog, *A prospecitve population study of psychosocial risk factors for late onset dementia.* International Journal of Geriatric Psychiatry, 1996. 11(1): p. 15-22

7. R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.