



Prospects of Statistical and Biostatistical Techniques in the Study of Diagnosis, Survival Analysis, and Disease Progression of Alzheimer's Disease



Vishal Deo*

Department of Statistics, University of Delhi, India

Submission: February 20, 2018; Published: July 02, 2018

*Corresponding author: Vishal Deo, Department of Statistics, Ramjas College, University of Delhi, India, Email: vishaaldeo@gmail.com

Abstract

Finding potent and clinically standardized biomarkers is of utmost importance for early detection of any disease which further helps in developing prevention and cure techniques. Use of statistical techniques becomes inevitable to validate the predictive power of the biomarkers and also to ensure the reproducibility of results. Identification of statistically significant covariates and competing risks is another important aspect of any biostatistical study for understanding disease progression. This paper attempts to review some useful statistical and biostatistical techniques being utilized in the study of dementia, Alzheimer's disease (AD) in particular, and also proposes some potentially reliable techniques which can be helpful in critically examining the complexities involved in brain activities of AD patients and its association with other pathological and physical changes in the patients. This paper concludes by proposing an objective-wise methodological structure for biostatistical analyses of AD.

Keywords: Alzheimer's Disease; Biomarkers; Combination of biomarkers; Disease progression; Survival Analysis; Brain Classification; Dementia; Competing risks; Cerebrospinal fluid; Scale-cognitive; Accelerated failure; Cardio vascular diseases; Dementia; Calculate hazards; Neuroimaging; Noninvasive techniques; Amyloid; Biochemical; Atrophy; Sensory-motor cortex; Spectroscopy; Antioxidant glutathione

Abbreviations: AD: Alzheimer's Disease; CSF: Cerebrospinal Fluid; PET: Positron Emission Tomography; FDG: Fluorodeoxyglucose; MRS: MR Spectroscopy; ADAS-Cog: Assessment Scale-Cognitive Subscale; AFTM: Accelerated Failure Time Models; MCI: Mild Cognitive Impairment; VSP: Visuo Spatial Perception; fMRI: Functional MRI; MMSE: Mini Mental State Examination; CVD: Cardio Vascular Diseases; AFTM: Accelerated Failure Time Models

Introduction

Dementia, in any form, is a dreadful disease and its severe impact on the most vulnerable section of our society demands for a dedicated attention from the scientific sphere. Till date, most of the research work related to dementia has focused on the prevalence of the disease and its diagnosis. Various methodologies have been proposed to identify biomarkers which can efficiently diagnose the presence of dementia. However, the subtlety of differences in the symptoms and biomarker levels, for different types and different states of dementia, makes it a difficult task to arrive at an objective solution to the problem of diagnosis. Alzheimer's disease (AD), being the most common form of dementia prevalent throughout the world, will be the focus of this review work. Nearly 36 million people worldwide have Alzheimer's or a related dementia and another major concern comes from the fact that in only 1-in-4 of these people AD has been diagnosed. A detailed review study by Prince, et al.[1] has shown that this number is expected to almost double in every 20 years and will reach around 115.4 million by the year 2050. According to a special report of the Alzheimer's association

of USA, number of prevalent cases of AD among the Americans is estimated to be around 5.2 million which includes 5 million people of age 65 and above and 200,000 people of age below 65. To put it in a nutshell, dementia (AD in particular) is going to be a huge burden on our society unless we try our best and determine efficient techniques for its diagnosis, prevention and cure.

Literature review and study prospects

Biostatistical study of a disease is imperative to establish scientific understanding of the disease and develop effective treatment regimes. The complexities of brain regions affected in dementia makes the use of statistical techniques crucial to identify biomarkers and validate their efficiency, map the affected brain regions, identify competing risk factors, calculate hazards and survival times, and most importantly to study the disease progression.

Diagnosis (Biomarkers and Brain classification)

Any statistical study of a disease requires extensive knowledge of the significant covariates or biomarkers associated

with the disease. Biomarkers with high predictive accuracy enable objective and timely diagnosis of the disease. Biomarkers being studied in AD research can be broadly classified into two types, invasive and noninvasive. Biomarkers based on blood samples and Cerebrospinal fluid (CSF) are invasive, while various kinds of neuroimaging biomarkers are based on noninvasive techniques.

Blood-based Biomarkers: Investigators have been working on developing blood based bio-markers for AD which will be less-costly and easily available but no standardized conclusion has been reached yet. The failure to replicate findings is the biggest obstacle in establishing credibility of blood-based biomarkers [2]. Biomarkers like Amyloid- β ($A\beta$), the ratio of $A\beta$ 1-42 to $A\beta$ 1-40 [3], and $A\beta$ 1-17 Perez, et al. [4] have been found to have potential of becoming good predictors of MCI and AD. The lack of statistical reproducibility of results across different cohorts may be attributed to factors like age, underlying cognitive health of subjects Toledo et al. [5], the diet of the subjects and other between subjects variability. Proper statistical frailty models are required to study the predictive capability of these biomarkers in the presence of such between subjects variability.

Cerebrospinal fluid (CSF) biomarkers

Presence of CSF in the extracellular space of the brain makes it a potentially good biomarker as it provides information about the biochemical changes undergoing in the brain regions. An increase in the levels of CSF tau and ptau (~300%) and decrease in $A\beta$ 42 (~50%) act as a good indicator of AD with sensitivity and specificity more than 80% [6]. A combined measure of these biomarkers (CSF tau, ptau and CSF $A\beta$ 42) performs better than each biomarker individually [7,8]. Changes in the levels of these biomarkers have been successfully used to study disease progression and identify the continuum phases of AD [9,10].

Neuro imaging biomarkers

Structural MRI, functional MRI and molecular imaging (MRS and PET) are the most common neuroimaging techniques being used to discover reliable biomarkers of MCI and AD. Structural MRI focuses on identifying brain regions with significant amount of atrophy in AD patients. A pattern of atrophy in the medial and lateral temporal lobe, medial and lateral parietal lobe and the frontal lobe, with visibly insignificant atrophy in the occipital lobe and sensory-motor cortex has been confirmed by various studies [11-13]. Functional MRI (fMRI) is used to assess the level of brain activity during a cognitive, sensory, or motor test, or at rest, by measuring blood flow and blood oxygen levels in specific regions. Recent studies have shown that changes in Visuospatial Perception (VSP) functions can be detected in early stages of AD using fMRI and the changes in VSP functions has been deemed to be a promising biomarker of AD [14].

Positron Emission Tomography (PET) uses radiolabelled ligands to measure metabolic and neurochemical processes in vivo. In general, AD research involves two types of PET ligands,

- i. Fluorodeoxyglucose (FDG)- used for measuring brain mechanism, and
- ii. Amyloid tracers: which bind to fibrillar amyloid plaques [15].

Various FDG PET studies have successfully mapped significant changes in metabolism in specific brain regions in AD patients which are being reckoned as very useful biomarkers and predictors of AD.

MR Spectroscopy (MRS) is another efficient non-invasive technique to trace the neurochemical changes in specific brain regions due to AD and other forms of dementia. Importance of MRS in AD research is highlighted by the advent of invaluable predictive MRS biomarkers with high sensitivity and specificity in various recent studies [16]. In recent times, brain antioxidant glutathione (GSH) has emerged as a potent biomarker of MCI and AD. Mandal et al. [17] used in-vivo proton MRS to measure GSH levels in specific brain regions and studied the association between GSH levels and clinical measures of AD progression. The authors found statistical evidences of strong diagnostic potential of GSH, both for MCI and AD, and also unveiled the ability of GSH to distinguish within the continuum stages of AD with high accuracy.

Multivariate techniques like cluster analysis and discriminant analysis can also prove useful in assessing the effectiveness of the biomarkers in correct diagnosis and prediction of the disease. Guttula et al. [18] used hierarchical cluster analysis to identify biomarker genes for AD. Whitwell et al. [19] used hierarchical agglomerative cluster analysis using Ward's clustering linkage method to examine case-by-case variability in patterns of grey matter atrophy in subjects with the behavioural variant of frontotemporal dementia and were able to identify anatomical subtypes of frontotemporal dementia. Efforts have also been made to improve the performance of these biomarkers by combining their predictive capacities. Many authors have used logistic regression to find best linear combination of biomarkers with improved predictive accuracy and specificity. Pepe & Thompson [20] discussed the logistic regression technique to find a linear combination of biomarkers to optimize diagnostic accuracy which can maximize the area under the ROC curve. Shaffer et al. [21] conducted a study to identify the best combination of three biomarkers, MR imaging, FDG PET, and CSF proteins, which can, along with clinical information based on neuropsychological tests, most efficiently predict the conversion to AD from MCI (Mild Cognitive Impairment).

They constructed logistic regression models with dependent variable as "conversion within 4 years" and used age, education, ApoE genotype, and Alzheimer's Disease Assessment Scale-Cognitive subscale (ADAS-Cog) as covariates. Then they added the three biomarkers in the model sequentially to get every combination of variables in the logistic regression models and then compared the resultant models using the measures of the

area under the ROC curve, the Akaike information criterion, and the misclassification rate. Their study concluded that a model combining the three biomarkers with the clinical information was most accurate in predicting future MCI conversion to AD. Despite the fact that researchers have been able to find various potent biomarkers with good predictive value, the goal of standardization of diagnosis based on universally accepted biomarkers has not been achieved. We need to further deliberate on the current findings, and try to validate the existing results by testing their replicability for different cohorts. Also, we need to find combination of different kinds of biomarkers which can together give a much improved diagnostic and predictive accuracy. Since MCI and AD exhibit structural, functional as well as biochemical changes, a combination of biomarkers tracking all such changes may give desired result. Also, the theory behind logistic regression suggests that the estimates of the model maximize the logistic likelihood and is not directly linked to the maximization of discriminatory power of the linear combination of the biomarkers. This is another reason to be motivated to find

other more effective ways of combining the biomarkers based on multivariate techniques.

Methods based on logistic regression are very effective in constructing whole brain classifiers for fMRI data. Ryali et al. [22] and Krishnapuram et al. [23] have described techniques based on logistic regression to select relevant brain regions that discriminate between cognitive conditions which can be further used as brain classifiers. Voxels are taken as the input variables (independent variables) and class label (a binary variable) is taken as the dependent variable in the logistic regression model. The main problem faced by researchers in estimating the parameters of such models for brain classification arises from the fact that the number of voxels (independent variables) is far greater than the number of observations. To overcome this problem certain constraints are induced in the model which brings in the concept of sparse logistic regression (or logistic regression with regularization). The parameters of the model are estimated using the method of Maximum Likelihood Estimation (Figure 1).

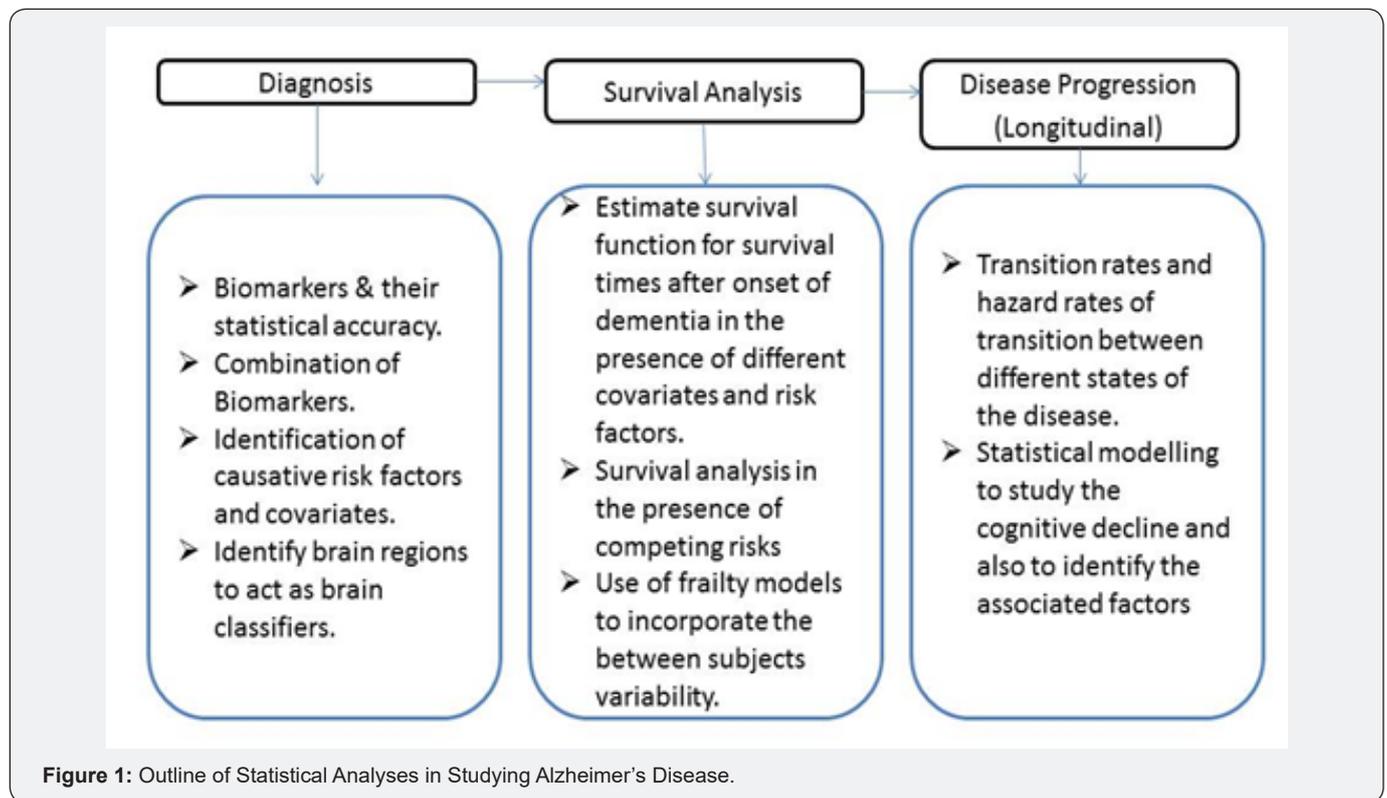


Figure 1: Outline of Statistical Analyses in Studying Alzheimer's Disease.

Survival analysis and disease progression

Understanding disease progression, both duration wise and intensity wise, is very crucial for the development of effective treatment and patient care techniques. Different non-parametric and parametric methods like Kaplan-Meier curves, Cox-proportional hazard models, Accelerated Failure Time Models (AFTM) etc. can be used for survival analysis. Xie et al. [24] used Kaplan-Meier curves and Cox's proportional hazard model for studying survival times in dementia patients and estimated

the survival function after onset of dementia in the presence of various covariates like age, sex, disability and severity of cognitive impairment. Magierski et al. [25] used Kaplan-Meier curves and Cox regression analysis to compare the disease progression (survival times) of dementia with lewy bodies and Alzheimer's disease.

A lot more informative study can be conducted by introducing functional, structural, and biological changes in the affected brain regions as covariates/biomarkers in these models. Other

covariates may include various test scores, like Mini Mental State Examination (MMSE) etc. Factors like smoking habit, drinking habit, and medical conditions like diabetes and Cardio Vascular Diseases (CVD) should be included in the survival models. The diverseness in the brain activity of different individuals (even

in the same cognitive state) should motivate the use of Frailty models as an improvement over other models in survival analysis as they can account for heterogeneity among the individuals due to some immeasurable reasons (Table 1).

Table 1:

Purpose		Statistical/Biostatistical Techniques
Diagnosis		
1	Identify efficient biomarkers	ROC analysis Logistic Regression
2	Combine biomarkers and validate their predictive power and brain region classification power	Logistic regression Sparse Logistic regression Cluster Analysis
Survival Analysis		
1	Estimate survival rate for onset of AD in MCI patients/ Estimate survival rate of death after onset of AD	Non-parametric methods like Kaplan-Meier product limit estimation Parametric survival distributions with underlying distributions like weibull, exponential, gamma etc. and Accelerated failure time models Semi parametric survival models like proportional and nonproportional Cox hazard models (with or without frailty). Joint Modelling approach.
2	Study the survival rates in the presence of competing risks	Competing risks survival models
Disease Progression		
1	Estimate hazard rates and transition probabilities of transition between different states of AD (defined on the basis of biomarkers and covariates)	Markov Analysis
2	Study cognitive decline	Different regression techniques like linear, generalized linear, including Bayesian regression in the presence of significant factors and covariates.

Competing risks should be included in the survival models to get unbiased estimates of factor effects on the event of interest. Chang et al. [26] fitted survival models in the presence of competing risk as death to study the effect of smoking on AD. They concluded that effect of smoking on AD differs between models that are and models that are not adjusted for competing risks. Again, only a handful of literature is available on the study of disease progression in AD patients and more in depth work is required in this direction. Harezlaket al. [27] came up with an illness death stochastic model, for longitudinal dementia data, with three states, non-diseased, diseased and dead and used markov process analysis to simultaneously estimate disease incidence and mortality rates.

A more relevant and revealing result can be obtained by including Mild Cognitive Impairment (MCI) state in the markov model. Detailed investigation of transition from MCI to AD is essential to gauge the progression of cognitive impairment in the affected patients. Timely treatment of the patients with early onset of MCI can save them from developing AD or at least slow down the progression of AD. Hazard rates and transition probabilities of transition between different states can be calculated based on the defined markov chain model.

Longitudinal study can be extended to Bayesian stochastic models to study disease progression and validity of different models can be tested using simulation techniques. Houtet al. [28] have provided Bayesian inference for a continuous-time 3-state illness-death Markov model to calculate life expectancies for Parkinson's patients in presence of different risk factors.

Outline of Statistical Analyses in Studying Alzheimer's Disease

Biostatistical study of Alzheimer's disease (or any other dementia type) can be broadly classified into the following three stages.

Summary of methodology with relevance and importance

Diagnosis: Statistical techniques like ROC analysis and logistic regression can be utilized to analyze predictive power of potential biomarkers of AD. Best available biomarkers based on MRI, fMRI, and MRS which are highly capable of measuring structural, functional and biochemical changes in the brain regions of AD patients can be considered to find best linear combination using logistic regression technique. The discriminatory power of this linear combination can be validated

using multivariate techniques like factor analysis, discriminant analysis and cluster analysis.

Survival analysis: Once the identification of efficient biomarkers or combination of biomarkers is done, one can proceed with survival analysis of the patients suffering from AD before studying the disease progression. The aim at this stage will be to estimate survival function (survival rate) using the data on time to the occurrence of the event of interest. Survival analysis can be performed to achieve the following goals

- i. To estimate survival rate for onset of AD (for time until onset of AD) in MCI patients, and
- ii. To estimate the survival rate of death after onset of AD. Non parametric methods like Kaplan-Meier, parametric survival models like weibull, exponential, gamma etc., and semiparametric methods like proportional and nonproportional Cox hazard models are noble methods to be used to model the survival times and the models will be compared to find the best fitted model.

These models will allow estimating survival function in the presence of covariates and at the same time they will allow identification of significant prognostic factors related to the survival times. Censoring should be introduced in the model to handle the censored data (which is more likely for any clinical study). If the data shows the presence of competing risks, competing risks survival models becomes imperative. To make the study more comprehensive and idealistic, frailty terms need to be introduced in these models to account for the heterogeneity among individuals due to unmeasured randomness. Joint modelling approach is another potent technique to study the association of longitudinal measures on biomarkers/covariates on the survival times [29].

Disease progression: Disease progression can be studied using markov chain stochastic models where different states of the markov chain are defined according to the extent of cognitive decline of the patients. Different continuum states of AD should be demarcated based on certain pathological markers or their combination (outcome of the first section of the study) and markov analysis can be performed to estimate hazard rates and transition probabilities of transition between the defined states of AD. These hazard rates and transition probabilities will provide immense knowledge to comprehend the progression of AD. Different regression models like linear, non-linear, generalized linear, Bayesian, ARIMA, etc. can be fitted (validity to be check first based on the data) to study the effect of different covariates on the rate of cognitive decline. All the models should be desirably tested for their fit, compared using statistical indicators, and validated using simulation techniques.

References

1. Prince M, Bryce R, Albanese E, Wimo A, Ribeiro W, et al. (2013) The global prevalence of dementia: A systematic review and metaanalysis. *Alzheimers Dement* 9(1): 63-75.
2. Watt AD, Perez KA, Rembach AR, Masters CL, Villemagne VL, et al. (2012) Variability in blood-based amyloid-beta assays: the need for consensus on pre-analytical processing. *J Alzheimers Dis* 30: 323-336.
3. Koyama A, Okereke OI, Yang T, Blacker D, Selkoe DJ, et al. (2012) Plasma amyloid-beta as a predictor of dementia and cognitive decline: a systematic review and meta-analysis. *Arch Neurol* 69(7): 824-831.
4. Pérez GV, Pesini P, Allué JA, Sarasa L, Montañés M. et al. (2012) Beta-amyloid-17 is a major beta-amyloid fragment isoform in cerebrospinal fluid and blood that shows diagnostic value. *J Alzheimers Dis* 8(4): 247.
5. Toledo JB, Vanderstichele H, Figurski M, Aisen PS, Petersen RC, et al. (2011) Factors affecting A beta plasma levels and their utility as biomarkers in ADNI. *Acta Neuropathol* 122(4): 401-413.
6. Blennow K, Hampel H, Weiner M, Zetterberg H (2010) Cerebrospinal fluid and plasma biomarkers in Alzheimer Disease. *Nat Rev Neurol* 6(3): 131-144.
7. Galasko D, Chang L, Motter R, Clark CM, Kaye J, et al. (1998) High cerebrospinal fluid tau and low amyloid Beta 42 levels in the clinical diagnosis of Alzheimer's disease and relation to apolipoprotein E genotype. *Arch Neurol* 55(7): 937-945.
8. Maddalena A, Papassotiropoulos A, Müller TB, Jung HH, Hegi T, et al. (2003) Biochemical diagnosis of Alzheimer disease by measuring the cerebrospinal fluid ratio of phosphorylated tau protein to beta-amyloid peptide42. *Arch Neurol* 60(9): 1202-1206.
9. Perrin RJ, Fagan AM, Holtzman DM (2009) Multimodal techniques for diagnosis and prognosis of Alzheimer's disease. *Nature* 461(7266): 916-922.
10. Jack CR, Bernstein MA, Borowski BJ, Gunter JL, Fox NC, et al. (2010) Update on the magnetic resonance imaging core of the Alzheimer's Disease neuroimaging initiative. *Alzheimers Dement* 6(3): 212-220.
11. Jack CR, Knopman DS, Jagust WJ, Shaw LM, Aisen PS, et al. (2010) Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurol* 9(1): 119-128.
12. Risacher SL, Saykin AJ, West JD, Shen L, Firpi HA, et al. (2009) Baseline MRI predictors of conversion from MCI to probable AD in the ADNI cohort. *Curr Alzheimer Res* 6(4): 347-361.
13. Weiner MW, Veitch DP, Aisen PS, Beckett LA, Cairns NJ, et al. (2011) The Alzheimer's disease neuroimaging initiative: a review of papers published since its inception. *Alzheimers Dement* 8(1 suppl): S1-6.
14. Mandal PK, Joshi J, Saharan S (2012) Visuospatial Perception: An emerging biomarker for Alzheimer's Disease. *J Alzheimers Dis* 31(3): 117-136.
15. Risacher SL, Saykin AJ (2013) Neuroimaging and other biomarkers for Alzheimer's disease: The changing landscape of early detection. *Annu Rev Clin Psychol* 9: 621-648.
16. Mandal PK (2014) Brain metabolic mapping with MRS: A potent noninvasive tool for clinical diagnosis of brain disorders. *AJNR Am J Neuroradiol* 35 (6 suppl): S1-S3.
17. Mandal PK, Saharan S, Tripathi M, Murari G (2015) Brain glutathione levels: A novel biomarker for Mild Cognitive Impairment and Alzheimer's Disease. *Biol Psychiatry* 78(10): 702-710.
18. Guttula SV, Allam A, Gumpeny RS (2012) Analyzing Microarray Data of Alzheimer's Using Cluster Analysis to identify the Biomarker Genes. *Int J Alzheimers Dis*.
19. Whitwell JL, Przybelski SA, Weigand SD, Ivnik RJ, Vemuri P, et al. (2009) Distinct anatomical subtypes of the behavioural variant of frontotemporal dementia: a cluster analysis study. *BRAIN* 132(Pt 11): 2932-2946.
20. Pepe MS, Thompson ML (2000) Combining diagnostic test results to increase accuracy. *Biostatistics* 1(2): 123-140.

21. Shaffer JL, Petrella JR, Sheldon FC, Choudhury KR, Calhoun VD, et al. (2013) Predictive cognitive decline in subjects at risk for Alzheimer's disease by using combined cerebrospinal fluid, MR Imaging, and PET biomarkers. *Radiology* 266(2): 583-591.
22. Ryali S, Supekar K, Abrams DA, Menon V (2010) Sparse logistic regression for whole brain classification of fMRI data. *Neuroimage* 51(2): 752-764.
23. Krishnapuram B, Carin L, Figueiredo MA, Hartemink AJ (2005) Sparse multinomial logistic regression: Fast algorithms and generalization bounds. *IEEE transactions on pattern analysis and machine intelligence* 27(6): 957-968.
24. Xie J, Brayne C, Matthews FE, Medical research council cognitive function and ageing study collaborators (2007) Survival times in people with dementia: analysis from population based cohort study with 14 year follow-up. *BMJ*.
25. Magierski R, Kłoszewska I, Sobów TM (2010) The influence of vascular risk factors on the survival rate of patients with dementia with Lewy bodies and Alzheimer disease. *Neurol Neurochir Pol* 44(2): 139-147.
26. Chang CC, Zhao Y, Lee CW, Ganguli M (2012) Smoking, death, and Alzheimer's disease: A case of competing risks. *Alzheimer Dis Assoc Disord* 26(4): 300-306.
27. Harezlak J, Gao S, Hui SL (2003) An illness-death stochastic model in the analysis of longitudinal dementia data. *Stat Med* 22(9): 1465-1475.
28. Van den Hout A, Matthews FE (2009) Estimating dementia-free life expectancy for Parkinson's patients using Bayesian inference and microsimulation. *Biostatistics* 10(4): 729-743.
29. Grover G, Prafulla K, Vishal Deo, Manoj K V (2015) A Joint Modelling Approach to Assess the Impact of CD4 Cell Count on the Risk of Loss to Follow up in HIV/AIDS Patients on Antiretroviral Therapy. *International Journal of Statistics and Applications* 5(3): 99-108.



This work is licensed under Creative Commons Attribution 4.0 License
DOI: [10.19080/BBOJ.2018.07.555712](https://doi.org/10.19080/BBOJ.2018.07.555712)

Your next submission with Juniper Publishers will reach you the below assets

- Quality Editorial service
- Swift Peer Review
- Reprints availability
- E-prints Service
- Manuscript Podcast for convenient understanding
- Global attainment for your research
- Manuscript accessibility in different formats
(Pdf, E-pub, Full Text, Audio)
- Unceasing customer service

Track the below URL for one-step submission
<https://juniperpublishers.com/online-submission.php>