BUILDING PREDICTION MODELS FOR DEMENTIA

THE NEED TO ACCOUNT FOR INTERVAL CENSORING AND THE COMPETING RISK OF DEATH

Arika L. Marchetti

Submitted to the faculty of the University Graduate School

in partial fulfillment of the requirements

for the degree

Master of Science

in the Department of Biostatistics,

Indiana University

August 2019

Accepted by the Graduate Faculty of Indiana University, in partial fulfillment of the requirements for the degree of Master of Science.

Master's Thesis Committee

_____
Giorgos Bakoyannis, PhD, Chair

_____
Xiaochun Li, PhD

_____
Sujuan Gao, PhD

_____
Constantin Yiannoutsos, PhD

ACKNOWLEDGEMENT

I would like to thank my thesis advisor Dr. Giorgos Bakoyannis of the Department of Biostatistics at Indiana University. His office door was always open when I ran into questions or needed an equation or model explained.

I would also like to thank the members of my committee for this research project: Dr. Sujuan Gao, Dr. Constantin Yiannoutsos, and Dr. Xiaochun Li.

Arika L. Marchetti

BUILDING PREDICTION MODELS FOR DEMENTIA THE NEED TO ACCOUNT FOR INTERVAL CENSORING AND THE COMPETING RISK OF DEATH

Context. Prediction models for dementia are crucial for informing clinical decision making in older adults. Previous models have used genotype and age to obtain risk scores to determine risk of Alzheimer's Disease, one of the most common forms of dementia (Desikan et al., 2017). However, previous prediction models do not account for the fact that the time to dementia onset is unknown, lying between the last negative and the first positive dementia diagnosis time (interval censoring). Instead, these models use time to diagnosis, which is greater than or equal to the true dementia onset time. Furthermore, these models do not account for the competing risk of death which is quite frequent among elder adults.

Objectives. To develop a prediction model for dementia that accounts for interval censoring and the competing risk of death. To compare the predictions from this model with the predictions from a naïve analysis that ignores interval censoring and the competing risk of death.

Methods. We apply the semiparametric sieve maximum likelihood (SML) approach to simultaneously model the cumulative incidence function (CIF) of dementia and death while accounting for interval censoring (Bakoyannis, Yu, & Yiannoutsos, 2017). The SML is implemented using the R package **intccr**. The CIF curves of dementia are compared for the SML and the naïve approach using a dataset from the Indianapolis Ibadan Dementia Project.

Results. The CIF from the SML and the naïve approach illustrated that for healthier individuals at baseline, the naïve approach underestimated the incidence of dementia compared to the SML, as a result of interval censoring. Individuals with a poorer health condition at baseline have a CIF that appears to be overestimated in the naïve approach. This is due to older individuals with poor health conditions having an elevated risk of death.

Conclusions. The SML method that accounts for the competing risk of death along with interval censoring should be used for fitting prediction/prognostic models of dementia to inform clinical decision making in older adults. Without controlling for the competing risk of death and interval censoring, the current models can provide invalid predictions of the CIF of dementia.

Giorgos Bakoyannis, PhD, Chair

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter One Introduction

Dementia

Dementia is the loss of cognitive functioning and behaviors that negatively impacts an individual's daily life (National Institute on Aging, 2017b). The World Health Organization (WHO) estimates around 50 million people have some form of dementia worldwide (World Health Organization, 2019). There are four major types of dementia, and it is common for people to have mixed dementia, which is a combination of two or more types of dementia. These types include Lewy Body dementia, frontotemporal disorders, vascular dementia, and Alzheimer's disease (National Institute on Aging, 2017b). Alzheimer's disease is characterized by amyloid plaques and neurofibrillary tangles in the brain, as well as the loss of connections between neurons. Alzheimer's Disease is the most common form of dementia in older adults, and it affects approximately 5.5 million people in the US over the age of 65, and it is the sixth leading cause of death in the US (National Institute on Aging, 2017a).

Risk Factors

Dementia has been found to be associated with non-insulin-dependent diabetes mellitus, also known as Type II diabetes (A. Ott et al., 1996). In addition to diabetes as a potential risk factor for Alzheimer's disease, there are multiple other risk factors that could be included that are associated with diabetes. Excess body fat has been shown to be associated with increased risk of Type-2 diabetes in women (Hu et al., 2001). One study used a GSMR method that performed a randomization analysis to test for associations between common health risk factors. This study found that LDL-cholesterol has a protective effect against type-2 diabetes, the number of years of education has a

protective effect against Alzheimer's disease, and an increase in BMI increases the risk of type 2 diabetes, but Type-2 diabetes reduces the risk of high BMI (Zhu et al., 2018). While diabetes has been found to be associated with a cognitive decline, interventions before the age of 60 have been found to lessen this cognitive decline. This was assessed using multivariate analysis while controlling for demographics to compare scores on the cognitive tests. The change scores were controlled for risk factors such as age, gender, center, educational level, site, and CNS-relevant medications using linear regression (Knopman et al., 2001).

Not only has diabetes been associated with Alzheimer's disease and related syndromes, or ADRS, but Alzheimer's can also influence diabetes as well. ADRS was found to be associated with a decrease in monitoring of diabetes and therefore an increase in complications due to diabetes. The study compared groups with and without ADRS using standardized incidence ratios. Data of censored records was used up until the time of censoring (Wargny et al., 2018). This study suggests that Alzheimer's disease can have other impacts on the health of individuals, possibly creating a cycle of health problems.

Age has an effect on dementia as well. Older ages have been associated with higher risk. Gender also has an effect on dementia; similar incidence rates of dementia were found in men and women up to age 85, but the lifetime risk of dementia was found to be higher in women than in men. This also reflects the fact that women have a higher life expectancy than men (Alewijn Ott, Breteler, Harskamp, Stijnen, & Hofman, 1998). Low level of education was also found to be significantly associated with dementia (Zhang et al., 1990).

Prediction/Prognostic Models

Prognostic models can be used to predict the risk of an event occurring in the future, such as developing a disease (Cook, 2008). Prognostic models is a term used to describe clinical prediction models. Prediction models of a certain prognosis are useful for physicians to make decisions about screening and treatment or therapy for individuals that are at a high risk of developing the disease. Clinical prediction models can provide evidence of risks in order for the physician and individual to work together for decision making as well (Steyerberg, 2008). To be useful in a clinical setting, these models need to be valid methodologically (Perel, Edwards, Wentz, & Roberts, 2006). Predictive modeling performs a risk assessment to identify individuals that are at risk for a specific disease, which allows time for action. They can be used to determine the best clinical choices for the patient in a specific situation, as well as used in decisions about the overall healthcare system (Vogenberg, 2009).

Decline in cognitive functions has been found to occur years before dementia can be clinically diagnosed (Amieva et al., 2005). The importance of studying preventive treatments administered before the onset of Alzheimer's symptoms for individuals at high risk of developing Alzheimer's is of scientific interest as well (Aisen et al., 2011). Epidemiological studies are being used to determine which risk factors have effects on the outcome of dementia, and the risks and benefits of various prevention strategies. This information can then be relayed to physicians (Patterson et al., 2008). Potential treatment options include behavioral modifications and both non-pharmacological and pharmacological interventions to manage symptoms (Mayo Clinic, 2019). However, a review of some pharmacological therapies has suggested they may not be as effective in

managing symptoms of dementia (Sink, Holden, & Yaffe, 2005). Further research is needed to determine associations between risk factors and dementia in order to individualize treatment based on known present risk factors, and prognostic models may be useful in this aspect.

Previous prediction models for Alzheimer's Disease do not account for the fact that the time to dementia onset is unknown, resulting in interval censoring. Instead, these models use time to diagnosis, which is greater than or equal to the true dementia onset time. Furthermore, these models do not account for the competing risk of death which is quite frequent among elder adults. Previous models have used genotype and age to obtain risk scores to determine risk of Alzheimer's Disease, one of the most common forms of dementia. The age of Alzheimer's onset was taken at the time when symptoms began. (Desikan et al., 2017). The landmark Rotterdam Study used the midpoint between baseline age and age at diagnosis (Alewijn Ott et al., 1998). Accuracy of predictive tools have been assessed using area under the Receiver Operating Characteristic (AUC) and Brier Scores (BS) for data with competing risks. However, this study also used the mid-point between diagnosis and the time of last visit without dementia (Blanche et al., 2015).

Previous models have accounted for competing risks in various ways, but few studies have looked at interval censoring and the competing risk of death in dementia prognostic models. Competing risks have been found to be critical in assessing the risk of disease for older individuals. Standard univariate survival analysis without competing risks was found to overestimate risk of hip fracture when compared to estimates accounting for competing risks (Berry, Ngo, Samelson, & Kiel, 2010). Smoking, death, and Alzheimer's Disease were assessed with competing risk of death due to the association between

smoking and death. Failure to account for this competing risk of death resulted in biased results (Chang, Zhao, Lee, & Ganguli, 2012).

Chapter Two Methodology

Survival Analysis

Survival analysis is a method of analyzing data to assess the time to an event of interest. In the health-related field, this event is commonly diagnosis of a disease or death. The survival probability is the probability that an individual survives from the start of the study or event to a specific time. Common approaches to survival analysis include the nonparametric Kaplan-Meier estimator and the semiparametric Cox proportional hazards model. Data in survival analysis may be censored in different ways. Examples of censoring include the fact that not all individuals will have experienced the event during the study time, patients may be lost to follow-up for part of the study or the rest of the study, or a competing event may occur (Clark, Bradburn, Love, & Altman, 2003).

Competing Risks and Interval Censoring

Competing risks are alternative events that could occur which would prevent the observation of the event of interest. One of the major types of competing risks is death. If death should occur, it is unknown whether that individual would continue to live to the end of the study without experiencing the event of interest, or if the individual would develop the event of interest. Should these competing events be considered censored data, this may create bias in certain estimates, such as the Kaplan-Meier estimator (Putter, Fiocco, & Geskus, 2007). Even when the Kaplan Meier accounts for competing risk, the results are biased by tending to overestimate event rates, and the "cumulative incidence competing risks" method should be used instead. Kaplan-Meier approaches to competing risk data include the Kaplan-Meier censor all method, in which individuals are considered censored if a competing risk event occurs, the Kaplan-Meier censor death

only method, in which individuals are only censored if they experience death before the event of interest or are lost to follow-up while other competing risks are ignored, and the Kaplan-Meier ignore all method, which ignores all competing risks and only individuals lost to follow-up are censored. The Kaplan-Meier methods rely on the assumption that the censoring and the event of interest are independent, whereas the cumulative incidence competing risks method does not (Southern et al., 2006; Tai, Machin, White, & Gebski, 2001).

When addressing data with competing risks, there are two quantities of interest. The Cause-Specific Hazard reflects the instantaneous failure rate from a specific event in the presence of all other events. The Fine-Gray model, a more simple and straightforward approach than the Cumulative Specific Proportional Hazards model, creates a survival model based on the cumulative incidence function (Fine & Gray, 1999). The Cumulative Incidence Function (CIF), which is the cumulative probability of an event in the presence of all other events, is commonly used in prediction and prognosis models as it quantifies the absolute risk of an over time (Bakoyannis et al., 2017). More precisely, the CIF is defined as:

$$F_j(t) = \Pr(T \leq t, C = j) = \int_0^t \lambda_j(u) exp\left[- \int_0^u \sum_{c=1}^{k} \lambda_j(w) dw \right] du$$

Interval censoring occurs when the event of interest occurs between two observation times, such as two clinical visits (Putter et al., 2007). The standard Cox model does not account for interval censoring (Leffondré, Touraine, Helmer, & Joly, 2013). Previous “illness-death" models used a semi-parametric approach to model interval-censored data

and the estimates were found to be better than those of a more naïve survival model, especially when there were high mortality rates present that tended to increase the relative bias in other models. However, the interval censoring in this study differed in the interval that was observed. This study looked at the probability that an individual developed the disease between the last visit time and the time of death (Leffondré et al., 2013).

Estimation Methodology

A method to analyze interval-censored competing risks data was proposed by Bakoyannis, Yu, and Yiannoutsos (Bakoyannis et al., 2017). This method is based on a B-spline sieve maximum likelihood (SML) and estimates a class of semiparametric generalized odds rate transformation models for the cause-specific CIF.

The likelihood function accounts for interval censoring by including the interval in the equation, using the last observation time prior to the event (V) and the first observation time after the event (U) as well as unique indicator functions for the interval censoring and left censoring. The unknown parameters to be estimated are $\theta = (\theta'_1, \theta'_2)$, and the vector of the covariates of interest is Z. The indicator function δ_{ij} is used to express which individual (i) experienced which event (j) when the event time is interval-censored, while the indicator function δ_{ij}^{1} indicates when the time is left-censored (Park, Bakoyannis, & Yiannoutsos, 2019).

The likelihood function for interval-censored data is therefore

$$L(\theta; D) \propto \prod_{i=1}^{n} \left(\left\{ \prod_{j=1}^{k} [F_j(U_i; \boldsymbol{Z}_i, \boldsymbol{\theta}_j) - F_j(V_i; \boldsymbol{Z}_i, \boldsymbol{\theta}_j)]^{\delta_{ij}} \right\} \times \left\{ \prod_{j=1}^{k} [F_j(U_i; \boldsymbol{Z}_i, \boldsymbol{\theta}_j)]^{\delta_{ij}^1} \right\} \left[1 - \sum_{j=1}^{k} F_j(V_i; \boldsymbol{Z}_i, \boldsymbol{\theta}_j) \right]^{1-\delta_i} \right)$$

Semiparametric models for the CIFs are considered in the aforementioned methodology. These models contain both non-parametric and parametric parts. The nonparametric part is approximation using B-splines. A general class of semiparametric models for the CIF is the class of semiparametric transformation models which has the form

$$g_j[F_j(t; \boldsymbol{z})] = \varphi_j(t) + \beta_j^T \boldsymbol{z}$$

The method by Bakoyannis et al. estimates a special subset of the class of semiparametric transformation models, known as the Generalized Odds-Rate Transformation models (Bakoyannis et al., 2017).

$$g_j(F_j; \alpha_j) = \begin{cases} log\left[\frac{(1-F_j)^{-\alpha_j}-1}{\alpha_j}\right], & \text{if } 0 < \alpha_j < \infty \\ log[-log(1-F_j)], & \text{if } \alpha_j = 0 \end{cases}$$

The Fine-Gray model is a specific member of the class of Generalized Odds-Rate Transformation models in which the alpha value equals zero, while the proportional odds model has an alpha value equal to one (Jeong & Fine, 2006; Park et al., 2019).

The semiparametric SML approach is used to estimate the parameters of these models under the likelihood defined above. This method produces regression coefficient estimates that are consistent, asymptotically normal, and semi-parametrically efficient

(Bakoyannis et al., 2017). In this work we use the above estimation methodology to develop a prediction model for dementia that accounts for interval censoring and the competing risk of death. Then, predictions from this model are compared to naïve predictions that ignore interval censoring and the competing risk of death by using a Cox hazards model.

Intccr R Package

The **intccr** package uses the function **ciregic** to fit the semiparametric regression model for the CIF that is in the class of generalized odds rate transformation models for this type of dataset that contains interval censoring and competing risk of death. It does this by assigning values to the competing event types (death or dementia) and identifying the right censored observations. The package utilizes parallel computing for a more efficient bootstrap variance-covariance matrix (Park et al., 2019).

Chapter Three Results

Participant Sample

The data used in the analysis came from a longitudinal study called the Indianapolis-Ibadan Dementia Project. The project contained data on two different populations, elderly African Americans from Indianapolis and elderly Yoruba from Ibadan, Nigeria. Our subset of data was from the Indianapolis cohort, and contained information on various potential risk factors for dementia. Participants were also classified as being in the new or old cohort depending on when they were enrolled in the study. Cohorts were enrolled in 1992 and in 2001 (Murray et al., 2018).

Statistical Analysis

First, we created a baseline dataset to identify status of each risk factor at baseline as well as the right censoring events to identify if the individual survived without dementia, developed dementia, or died as well as the interval in which the event occurred in terms of their age. Individuals that died before the age of 65 were removed as we wanted to only look at individuals surviving to the age of 65 which were dementia-free at that age. The **intccr** R package function **ciregic** (Park et al., 2019) was then applied to fit semiparametric models for the CIF using the B-spline SML approach by Bakoyannis et al. (2017).

We performed a model selection and selected the Fine-Gray model for dementia and the proportional odds model for death because they produced the greatest log-likelihood. A backwards selection of variables was performed to determine the variables that should be included in the Fine-Gray model based on a p-value cutoff of 0.05. The variables selected in backwards selection for dementia were baseline depression, baseline

heart attack status, and baseline stroke, while the variables selected for death were sex, baseline diabetes, baseline alcoholism, baseline depression, baseline heart attack status, baseline stroke status, and baseline smoking status.

The Fine-Gray model of the CIF was performed for dementia and death individually, and the estimates can be found in Table 1.

Table 1. Fine-Gray Model of Dementia Results

Parameter	Estimate	Standard Error	P value	Hazard Ratio
Sex (F)	-0.00705	0.10852	0.9482	0.993
Baseline Diabetes	-0.30712	0.11476	0.0074	0.736
Baseline Alcohol	-0.12256	0.11001	0.2652	0.885
Baseline Depressed	0.29304	0.15396	0.0570	1.341
Baseline Stroke	0.22833	0.13067	0.0806	1.257
Baseline Smoking	-0.39396	0.10253	0.0001	0.674

Statistical Analysis Results

The total number of individuals in the dataset was 4,095. 34.55% of the individuals did not develop dementia, 11.01% developed dementia, and 54.43% died before the end of the study (Table 2).

Table 2. Baseline Descriptive Statistics of Selected Variables for Participants without Dementia, with Dementia, and Participants that Died.

		Final Dementia status			Total Individuals
		No Dementia	Dementia	Died	
		N (%)	N(%)	N(%)	N(%)
Total Individuals		1415 (34.55)	451 (11.01)	2229 (54.43)	4095 (100)
Baseline Diabetes status	No Diabetes	1097 (77.53)	345 (76.50)	1524 (68.37)	2966 (72.43)
	Diabetes	318 (22.47)	106 (23.50)	705 (31.63)	1129 (27.57)
Gender	Female	1022 (72.23)	305 (67.63)	1333 (59.80)	2660 (64.96)
	Male	393 (27.77)	146 (32.37)	896 (40.20)	1435 (35.04)
Cohort	New	966 (68.27)	151 (33.48)	767 (34.41)	1884 (46.01)
	Old	449 (31.73)	300 (66.52)	1462 (65.59)	2211 (53.99)
		Final Dementia status			**Total Individuals**
		No Dementia	Dementia	Died	
		Median [IQR]	Median [IQR]	Median [IQR]	Median [IQR]
Years of education		12.00 [10.00,13.00]	9.00 [8.00, 12.00]	11.00 [8.00, 12.00]	11.00 [8.00, 12.00]
Baseline Age		73.58 [70.87, 77.81]	78.08 [72.93, 83.06]	75.21 [70.82, 80.49]	74.92 [70.99, 79.96]

For the purpose of the plots of the CIF for both the naïve Cox Proportional Hazards models and the SML models, covariate patterns known to be associated with dementia were used to observe the changes in the curves. The CIFs of dementia were plotted against age for males and females for the SML model and the naïve analysis.

The CIF from the SML and the naïve approach illustrated that for healthier individuals at baseline, the naïve approach underestimated the incidence of dementia compared to the SML, as a result of interval censoring (Fig 1). Individuals with a poorer health condition at baseline have a CIF that appears to be overestimated in the naïve approach (Fig 2). This is due to the fact that older individuals with a poor health condition have an elevated risk of the competing risk of death. Looking at one baseline risk factor at a time, for example, diabetes, shows us that the naïve approach

underestimates the CIF until approximately age 80, after which it begins to overestimate the CIF compared to the SML approach (Fig 3). The initial underestimation is a result of interval censoring, whereas the final overestimation is result of the competing risk of death at older ages.

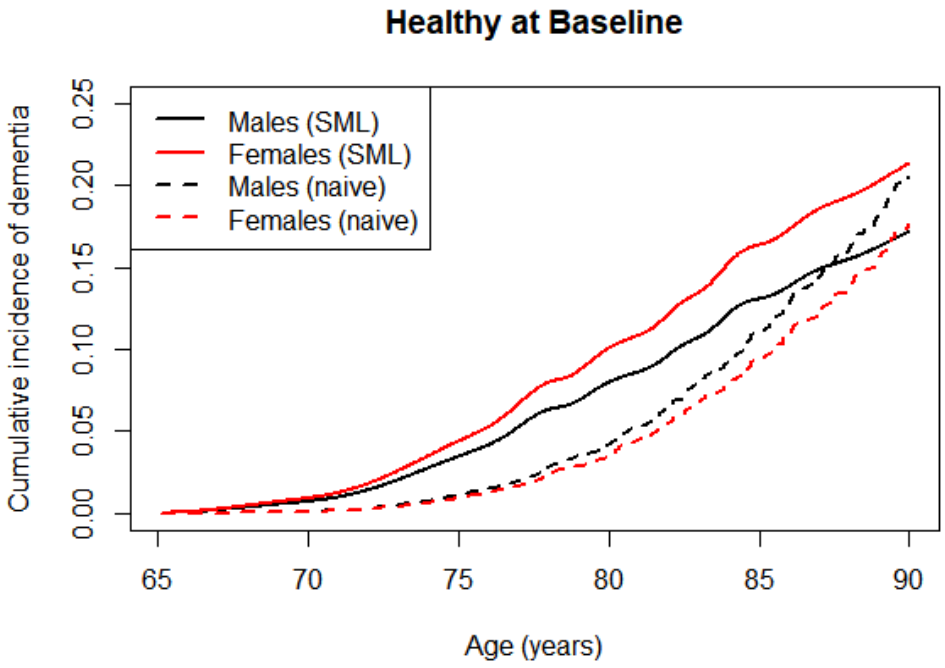

Figure 1. Cumulative Incidence Function of Dementia for Naïve Model and the Sieve Maximum Likelihood (SML) Model over Time Stratified by Gender for Individuals without Any Selected Risk Factors Present at Baseline.

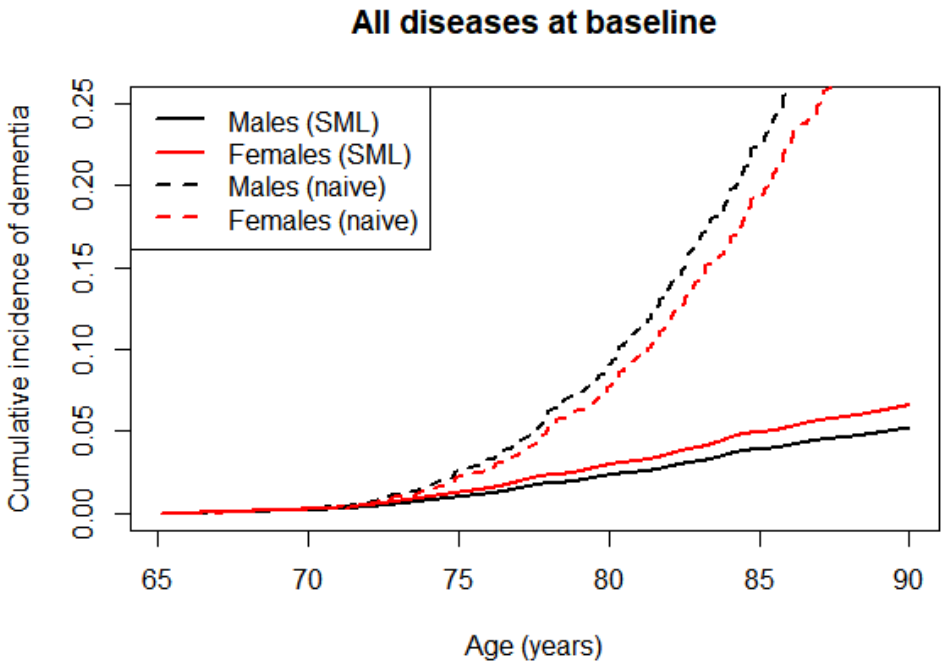

Figure 2. Cumulative Incidence Function of Dementia for Naïve Model and SML Model over Time Stratified by Gender for Individuals with All Selected Risk Factors Present at Baseline.

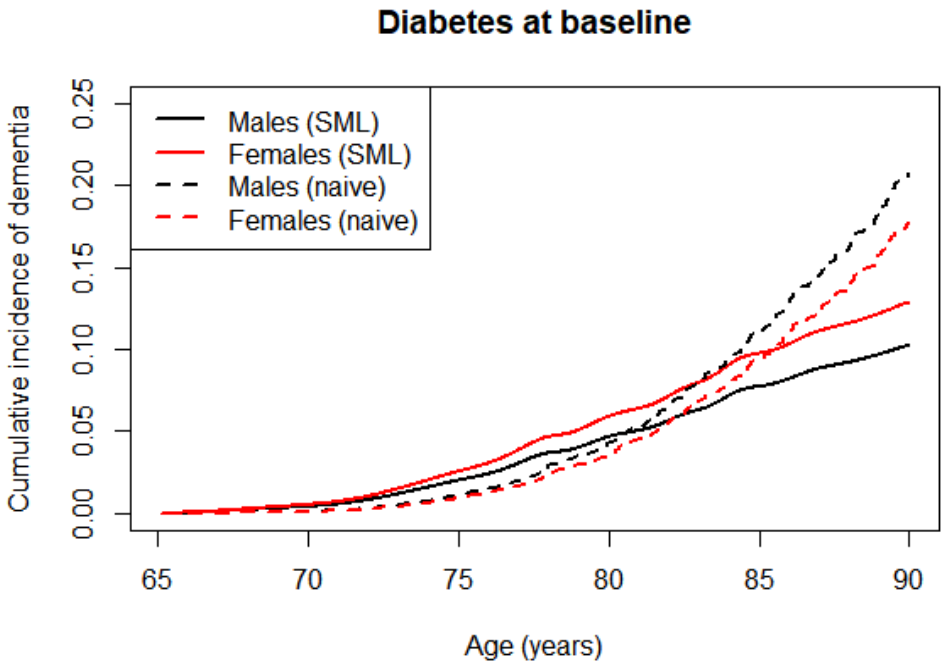

Figure 3. Cumulative Incidence Function of Dementia for Naïve Model and SML Model over Time Stratified by Gender for Individuals with Diabetes Present at Baseline.

Chapter Four Conclusions

While naïve approaches may be finding important and novel knowledge of associations between risk factors and diseases, they also yield biased results due to many prognostic models for dementia not accounting for interval censoring or competing risk of death. Interval censoring either puts the time of dementia onset at the midpoint between the last clinical visit or observed timepoint without diagnosis and the clinical visit or observed timepoint with diagnosis, or at the clinical visit or observed timepoint of diagnosis. The significant competing risk of death arises from the fact that dementia studies involve older adults with a higher mortality rate. This needs to be accounted for when building prediction models for dementia.

The **intccr** R package and **ciregic** function allow both interval censoring and competing risk to be controlled for by fitting a semiparametric SML approach to model the CIF of dementia while accounting for interval censoring. When applying this to prognostic factors in the Indianapolis Ibadan Dementia Project dataset, the CIF curves for the naïve analysis differ from our analysis depending on the age and the risk factors present.

In conclusion, the SML method that accounts for the competing risk of death along with interval censoring should be used for fitting prediction or prognostic models of dementia to inform clinical decision making in older adults. Without controlling for the competing risk of death and interval censoring, the current models can provide invalid predictions of the CIF of dementia. This indicates that accounting for interval censoring and the competing risk of death is necessary for dementia predictive modeling. While this

study does not consider time dependent covariates, future studies could address this as well.

References

Aisen, P. S., Andrieu, S., Sampaio, C., Carrillo, M., Khachaturian, Z. S., Dubois, B., . . . Vellas, B. (2011). Report of the task force on designing clinical trials in early (predementia) AD. *Neurology, 76*(3), 280-286. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/21178097 doi:10.1212/WNL.0b013e318207b1b9

Amieva, H., Jacqmin-Gadda, H., Orgogozo, J.-M., Le Carret, N., Helmer, C., Letenneur, L., . . . Dartigues, J.-F. (2005). The 9 year cognitive decline before dementia of the Alzheimer type: a prospective population-based study. *Brain, 128*(5), 1093-1101. Retrieved from https://doi.org/10.1093/brain/awh451. doi:10.1093/brain/awh451

Bakoyannis, G., Yu, M., & Yiannoutsos, C. T. (2017). Semiparametric regression on cumulative incidence function with interval-censored competing risks data. *Stat Med, 36*(23), 3683-3707. doi:10.1002/sim.7350

Berry, S. D., Ngo, L., Samelson, E. J., & Kiel, D. P. (2010). Competing Risk of Death: An Important Consideration in Studies of Older Adults. *Journal of the American Geriatrics Society, 58*(4), 783-787. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1532-5415.2010.02767.x. doi:10.1111/j.1532-5415.2010.02767.x

Blanche, P., Proust-Lima, C., Loubère, L., Berr, C., Dartigues, J.-F., & Jacqmin-Gadda, H. (2015). Quantifying and comparing dynamic predictive accuracy of joint models for longitudinal marker and time-to-event in presence of censoring and competing risks. *Biometrics, 71*(1), 102-113. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/biom.12232. doi:10.1111/biom.12232

Chang, C.-C. H., Zhao, Y., Lee, C.-W., & Ganguli, M. (2012). Smoking, death, and Alzheimer disease: a case of competing risks. *Alzheimer disease and associated disorders, 26*(4), 300-306. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/22185783 doi:10.1097/WAD.0b013e3182420b6e

Clark, T. G., Bradburn, M. J., Love, S. B., & Altman, D. G. (2003). Survival analysis part I: basic concepts and first analyses. *British journal of cancer, 89*(2), 232-238. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/12865907 doi:10.1038/sj.bjc.6601118

Cook, N. R. (2008). Statistical Evaluation of Prognostic versus Diagnostic Models: Beyond the ROC Curve. *Clinical Chemistry, 54*(1), 17. Retrieved from http://clinchem.aaccjnls.org/content/54/1/17.abstract. doi:10.1373/clinchem.2007.096529

Desikan, R. S., Fan, C. C., Wang, Y., Schork, A. J., Cabral, H. J., Cupples, L. A., . . . Dale, A. M. (2017). Genetic assessment of age-associated Alzheimer disease risk: Development and validation of a polygenic hazard score. *PLOS Medicine, 14*(3), e1002258. Retrieved from https://doi.org/10.1371/journal.pmed.1002258. doi:10.1371/journal.pmed.1002258

Fine, J. P., & Gray, R. J. (1999). A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association, 94*(446), 496-509. Retrieved from http://www.jstor.org/stable/2670170. doi:10.2307/2670170

Hu, F. B., Manson, J. E., Stampfer, M. J., Colditz, G., Liu, S., Solomon, C. G., & Willett, W. C. (2001). Diet, lifestyle, and the risk of type 2 diabetes mellitus in women. *N Engl J Med, 345*(11), 790-797. doi:10.1056/NEJMoa010492

Jeong, J.-H., & Fine, J. P. (2006). Parametric regression on cumulative incidence function. *Biostatistics, 8*(2), 184-196. Retrieved from https://doi.org/10.1093/biostatistics/kxj040. doi:10.1093/biostatistics/kxj040

Knopman, D., Boland, L. L., Mosley, T., Howard, G., Liao, D., Szklo, M., . . . Folsom, A. R. (2001). Cardiovascular risk factors and cognitive decline in middle-aged adults. *Neurology, 56*(1), 42-48. doi:10.1212/wnl.56.1.42

Leffondré, K., Touraine, C., Helmer, C., & Joly, P. (2013). Interval-censored time-to-event and competing risk with death: is the illness-death model more accurate than the Cox model? *International Journal of Epidemiology, 42*(4), 1177-1186. Retrieved from https://doi.org/10.1093/ije/dyt126. doi:10.1093/ije/dyt126

Mayo Clinic. (2019). Dementia. *Treatment*. Retrieved from https://www.mayoclinic.org/diseases-conditions/dementia/diagnosis-treatment/drc-20352019

Murray, M. D., Hendrie, H. C., Lane, K. A., Zheng, M., Ambuehl, R., Li, S., . . . Gao, S. (2018). Antihypertensive Medication and Dementia Risk in Older Adult African Americans with Hypertension: A Prospective Cohort Study. *Journal of General Internal Medicine, 33*(4), 455-462. Retrieved from https://doi.org/10.1007/s11606-017-4281-x. doi:10.1007/s11606-017-4281-x

National Institute on Aging. (2017a). What is Alzheimer's Disease? *Basics of Alzheimer's Disease and Dementia*. Retrieved from https://www.nia.nih.gov/health/what-alzheimers-disease

National Institute on Aging. (2017b). What is Dementia? *Basics of Alzheimer's Disease and Dementia*. Retrieved from https://www.nia.nih.gov/health/what-dementia

Ott, A., Breteler, M. M. B., Harskamp, F. v., Stijnen, T., & Hofman, A. (1998). Incidence and Risk of Dementia: The Rotterdam study. *American Journal of Epidemiology, 147*(6), 574-580. Retrieved from https://doi.org/10.1093/oxfordjournals.aje.a009489. doi:10.1093/oxfordjournals.aje.a009489

Ott, A., Stolk, R. P., Hofman, A., van Harskamp, F., Grobbee, D. E., & Breteler, M. M. B. (1996). Association of diabetes mellitus and dementia: The Rotterdam Study. *Diabetologia, 39*(11), 1392-1397. Retrieved from https://doi.org/10.1007/s001250050588. doi:10.1007/s001250050588

Park, J., Bakoyannis, G., & Yiannoutsos, C. T. (2019). Semiparametric competing risks regression under interval censoring using the R package intccr. *Computer Methods and Programs in Biomedicine, 173*, 167-176. Retrieved from http://www.sciencedirect.com/science/article/pii/S0169260718314615. doi:https://doi.org/10.1016/j.cmpb.2019.03.002

Patterson, C., Feightner, J. W., Garcia, A., Hsiung, G. Y. R., MacKnight, C., & Sadovnick, A. D. (2008). Diagnosis and treatment of dementia: 1. Risk assessment and primary prevention of Alzheimer disease. *Canadian Medical*

Association Journal, 178(5), 548. Retrieved from http://www.cmaj.ca/content/178/5/548.abstract. doi:10.1503/cmaj.070796

Perel, P., Edwards, P., Wentz, R., & Roberts, I. (2006). Systematic review of prognostic models in traumatic brain injury. *BMC Med Inform Decis Mak, 6*, 38. doi:10.1186/1472-6947-6-38

Putter, H., Fiocco, M., & Geskus, R. B. (2007). Tutorial in biostatistics: competing risks and multi-state models. *Stat Med, 26*(11), 2389-2430. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.2712. doi:10.1002/sim.2712

Sink, K. M., Holden, K. F., & Yaffe, K. (2005). Pharmacological Treatment of Neuropsychiatric Symptoms of DementiaA Review of the Evidence. *JAMA, 293*(5), 596-608. Retrieved from https://doi.org/10.1001/jama.293.5.596. doi:10.1001/jama.293.5.596

Southern, D. A., Faris, P. D., Brant, R., Galbraith, P. D., Norris, C. M., Knudtson, M. L., & Ghali, W. A. (2006). Kaplan–Meier methods yielded misleading results in competing risk scenarios. *Journal of Clinical Epidemiology, 59*(10), 1110-1114. Retrieved from http://www.sciencedirect.com/science/article/pii/S0895435606002630. doi:https://doi.org/10.1016/j.jclinepi.2006.07.002

Steyerberg, E. (2008). *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*: Springer Science & Business Media.

Tai, B. C., Machin, D., White, I., & Gebski, V. (2001). Competing risks analysis of patients with osteosarcoma: a comparison of four different approaches. *Stat Med, 20*(5), 661-684. doi:10.1002/sim.711

Vogenberg, F. R. (2009). Predictive and prognostic models: implications for healthcare decision-making in a modern recession. *American health & drug benefits, 2*(6), 218-222. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/25126292

Wargny, M., Gallini, A., Hanaire, H., Nourhashemi, F., Andrieu, S., & Gardette, V. (2018). Diabetes Care and Dementia Among Older Adults: A Nationwide 3-Year Longitudinal Study. *Journal of the American Medical Directors Association, 19*(7), 601-606.e602. Retrieved from http://www.sciencedirect.com/science/article/pii/S1525861017306849. doi:https://doi.org/10.1016/j.jamda.2017.12.006

World Health Organization. (2019). Dementia. Retrieved from https://www.who.int/news-room/fact-sheets/detail/dementia

Zhang, M., Katzman, R., Salmon, D., Jin, H., Cai, G., Wang, Z., . . . Liu, W. T. (1990). The prevalence of dementia and Alzheimer's disease in Shanghai, China: Impact of age, gender, and education. *Annals of Neurology, 27*(4), 428-437. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/ana.410270412. doi:10.1002/ana.410270412

Zhu, Z., Zheng, Z., Zhang, F., Wu, Y., Trzaskowski, M., Maier, R., . . . Yang, J. (2018). Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nature Communications, 9*(1), 224. Retrieved from https://doi.org/10.1038/s41467-017-02317-2. doi:10.1038/s41467-017-02317-2

CURRICULUM VITAE

Arika L. Marchetti

Education

- **Indiana University Fairbanks School of Public Health -** Indianapolis, IN
  Master of Science in Biostatistics, August 2019

- **University of Miami -** Coral Gables, FL
  Bachelor of Science in Neuroscience
  Minor in Chemistry, August 2015

Professional and Research Experience

- **Indiana University Fairbanks School of Public Health** – Indianapolis, IN
  Teaching Assistant, August 2018-December 2018

- **Indiana University Department of Environmental Health Science** – Indianapolis, IN
  Intern, May 2018-September 2018

- **University of Kansas Hospital Lean Promotion Office –** Kansas City, KS
  Intern, July 2017

- **University of Miami Neurology Department –** Coral Gables, FL
  Research Assistant, September 2015-March 2016

Organizations and Conferences

- **International Society for Clinical Biostatistics**
  Oral Presentation, July 2019

- **Indiana University Fairbanks School of Public Health Poster Session**
  Presented Poster, April 2019

- **Biostatistics Student Association**
  President, October 2017-August 2018

- **American Association for the Advancement of Science**
  Member and Volunteer, February 2011-Present

Awards

- **Service Learning Assistant scholarship**
  Sam H. Jones Community Service Scholarship Program, May 2018-August 2018