August 2013

Alcohol Biomarkers as Predictive Factors of Rearrest in High Risk Repeat Offense Drunk Drivers

Brian Charles Kay
University of Wisconsin-Milwaukee

ALCOHOL BIOMARKERS AS PREDICTIVE FACTORS OF REARREST IN HIGH RISK REPEAT OFFENSE DRUNK DRIVERS

by

Brian Kay

A Thesis Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Master of Science

in Health Care Informatics

at

The University of Wisconsin-Milwaukee

August 2013

ABSTRACT
ALCOHOL BIOMARKERS AS PREDICTIVE FACTORS OF REARREST IN HIGH RISK REPEAT OFFENSE DRUNK DRIVERS

by

Brian Kay

The University of Wisconsin- Milwaukee, 2013
Under the Supervision of Professor Rohit Kate

Alcohol biomarkers, or naturally occurring molecules which occur in response to one's alcohol consumption, are proving to be a value tool in objectively monitoring one's alcohol consumption. Coupling this assessment tool, with advances in computing power, new and powerful predictions are becoming evermore possible. In this retrospective study, data was first collected that consisted of a sample of 249 drivers convicted of driving under the influence charge and who monitored over the course of a year by biomarker blood tests. This data was then analyzed using machine learning methods. TwoStep cluster analysis showed distinct drinking groups within the drivers who were monitored. In addition to this, a cost sensitive learning classifier was utilized in order to predict if a driver would relapse, having a subsequent driving under the influence arrest. The algorithm was able to predict 64% of the cases within the training set. Additionally, learning curves indicated that correctly classified cases increased with the increase of training data, indicating that predictions may become more accurate with the availability of more training data.

Keywords: *alcohol, biomarkers, recidivism*

To my parents, for their love and support, and showing me all of the corners of the world, If it was not for you, I would not be who I am today; and to my love Michelle, without your love and support this would not be possible.

TABLE CONTENTS

LIST OF FIGURES

LIST OF TABLES

Introduction

Alcohol biomarkers are naturally occurring molecules which develop in response to the ingestion of alcohol. These molecules are proving to be an invaluable tool to objectively monitor the alcohol consumption for those deemed to be at a high risk of re-arrest for driving while intoxicated. This assessment tool combined with advances in computing power will allow for new and more powerful predictions in the future.

A survey released in 2009, by the United States Department of Health and Human Services, highlighted current rates of Driving While Under the Influence (DUI) and measured the way in which drunk driving is perceived by the survey participants. In the aforementioned article, twenty-six percent of Wisconsin adults stated that they had driven while intoxicated within the last year (SAMSHA, 2007). This reflects the highest occurrence of people driving while under the influence in the country; ranking Wisconsin as having the highest prevalence of drunk driving in the country.

In response to this study and in an effort to stop people from driving while intoxicated, preventative measures have been put into place to protect the public. These include an increased police presence throughout the community and various ad campaigns run through local media. While these interventions are designed to avert potential first time offenders; those deemed to be chronic

offenders present quite a different set of problems and challenges for treatment teams and increasing rates of recidivism.

1.1 Current approaches to reduce recidivism

Indeed it is clear that little has been done to prevent chronic offenders from endangering the public. Research has found that collision rates of these individuals are twice as high as the general population (Korzec, Bär, Koeter, & Kieviet, 2001). Drivers that have been convicted of more than three offenses are commonly classified as high-risk reoffenders or "hardcore drunk drivers." It is this specific population that presents two interesting challenges to both public law makers and alcohol assessment facilities. On one hand, assessment facilities are challenged with creating programs specifically designed to target repeat offenders; while on-the-other, community law makers must decide whether to increase penalties for those that continue to drive while intoxicated.

The well-being of the general population is put into jeopardy due to the specific "risks" this subgroup of drivers are willing to take. Measures must be taken to protect the public, as well as the "hardcore" offender. Research is finding that there are distinct causes of increased recidivism among those classified within this population. However, Alcohol Use Disorder appears to be one of the largest contributing factors (Couture, Brown, Tremblay, Kin, Ouimet, & Nadeau, 2012).

The assessment of Alcohol Use disorder has posed to be a difficult process through traditional methods. Research has proven that methods of alcohol consumption recall have been biased, where the majority of the bias is due to an individual underreporting or over reporting (Bean, Roska, Harasymiw, Pearson, Kay, & Louks, 2009). In many cases, the individual will underreport their use in an effort to guide treatment to their preference. As a result of these factors, new and novel approaches to curb the rising rates of recidivism as well as aides to more effectively diagnose Alcohol Use Disorder have been developed.

In Wisconsin, once the individual has been cited for driving under the influence, he/she can be referred to a state run assessment facility. The individual's alcohol and drug use will be assessed through an in-person interview. Based on the result of this assessment a driver's safety plan, or treatment protocol can then be established, whereby the person will be required to seek appropriate modalities of treatment.

1.2 Interlock Devices

One method to curb repeat offenders is the use of an interlock system. An interlock system is a device, when installed in a car, requires the driver to blow into a breathalyzer before the car can be started. If the interlock system detects that the driver has been drinking it will not allow the car to be started.

The first commercial interlocks, utilizing breathalyzers, were developed in the 1970's, however it was found that this system did not work particularly well

(Elder, et al., 2011). It was not until the 1990's that interlock systems became more widespread, effective, and ultimately more usable. With the advent of these " second generation" systems the National Highway Traffic Safety Administration began to fold them into individual drivers safety plans.

Today, it is estimated that 1.4 million drivers in the United States have an interlock system in their automobile (Elder, et al., 2011). However, this represents a small population in the United States and measures are being implemented in order to make interlocks more available. Based on the research of Elder (2011), it has been found that the interlocks utilized today are used in specific populations that are at a higher risk of offending, and those individuals that were offered the use of an interlock system to receive a reduced sentence.

An analysis of the data, for those drunk drivers likely to re-offend, discovered that the use of interlocks," Substantially, lower[s] the risk for recidivism than those who have had their licenses suspended either after being deemed ineligible for an interlock or deciding not to have one installed." (Elder, et al., 2011). Additionally, one study found that 93% of individuals who were deemed eligible for installing an interlocked had it done (Voas, Tippets , Fisher, & Grosz, 2010).

However, interlocks have a multitude of problems, which hamper their ability to become a widespread solution to curb drunk driving. Unfortunately, interlock devices are typically expensive to install and maintain. In the state of Wisconsin the cost of an interlock device can typically cost between $75-$150 to install;$60-$90 for monthly maintenance, leasing fees and removal fees which

can cost from $40-$60 (Wisconsin Department of Transportation, 2010). These devices require the offender to present to a specified facility in order for the device to be installed. Unfortunately, the devices are easily cheated. For example, an intoxicated driver can have a non-intoxicated passenger blow into the device in order to start the vehicle. Furthermore limited effectiveness rates have not been established on this treatment intervention.

1.3 Alcohol Biomarkers

Biological based indicators are also being utilized in the state of Wisconsin to monitor repeat offense drunk drivers. One of these approaches is monitoring high-risk subjects through the use of biomarker monitoring. A biomarker is,

> " A biological indicator that develops in the body when the person consumes alcohol, and stays elevated for longs periods of time- weeks, even months – after the person has stopped drinking" (Walker, 2012).

Biomarkers have been utilized in Europe for close to 30 years to monitor drivers who have been convicted of driving under the influence (Appenzeller, Schneider, Maul, & Wennig, 2005). This pioneering model monitors drunk drivers once they have been convicted, for a period of one year. During this time, if a biomarker result is positive then the driver will lose his/her driver's license indefinitely. If the driver successfully goes through this time without a positive biomarker then their driver's license will be reissued (Couture, Brown, Tremblay, Kin, Ouimet, & Nadeau, 2012). The increased level of monitoring allows the

driver to have proper allocation of resources for addressing their struggles with alcohol. If the biomarker shows that the individual's alcohol consumption is elevating, other interventions can be implemented in order to reduce this behavior.

1.4 Commonly Used Biomarkers

Currently, there are two general categories of alcohol biomarkers, those, which directly measure a derivative of the ethanol molecule in the body (direct), and those that measure the toxic effects of alcohol on one's system (indirect). The following include some of the biomarkers that are commonly used in the state of Wisconsin and also some that are used in the European model:

Carbohydrate-deficient Transferrin (CDT)

The gold standard in indirect biomarkers is Carbohydrate-deficient Transferrin (CDT). This molecule is a derivative of glycoprotein transferrin, which is the main molecule that carries iron in the body's bloodstream. When an individual consumes over 60g of alcohol per day, for over 2 weeks the body produces the Carbohydrate-deficient version on the molecule. CDT produces a high diagnostic specificity (The ability to indicate the absence of a CDT in a "truly" negative sample) for heavy alcohol use. Javors and Johnson (1998) established this rate at 93%. In addition, low incidences of false positives have been established due to CDT being highly specific to alcohol consumption (Peterson, 2004/2005). With the high specificity as well as low instances of false

positives, CDT proves to be an excellent marker if one is chronically heavy drinking.

Serum Gamma-Glutamyl Transferase (GGT)

The indirect molecule of *Serum Gamma-Glutamyl Transferase (GGT)* is a measure of liver function in the body. Physiologically, GGT is elevated when an individual consumes greater than 40g of alcohol per day for those deemed chronic alcoholics and 60g of alcohol per day in previous non-chronic alcoholics. GGT has varied sensitivity from 60 to 90%, documented in the research of Behrens et. al 1988. GGT has varied specificity, which is documented in the range of 55% to 100% (Sharpe, 2001).

GGT has a relatively long duration in the body, it remains detectable 14 to 26 days after one stops drinking and ceases to be elevated after 4-5 weeks. Unfortunately, GGT levels can be influenced by other illegal substances as well as legal drugs. Physiological disorders such as, obesity, diabetes, and clotting disorders can also influences levels of GGT in one's body (Rosman & Lieber, 1990).

Early Detection of Alcohol Consumption (EDAC)

The Early Detection of Alcohol Consumption test is a statistical analysis of multiple routine lab tests designed to detect high levels of alcohol consumption in the body. Utilizing Linear Discriminant Function, the EDAC produces a probability if an individual is regarded as a heavy drinker or an at-risk drinker.

The differentiation between these two types of drinkers allows clinicians to ascribe appropriate resources to the individuals.

In a general population the EDAC performed marginally, producing 30% sensitivity in males and 42% sensitivity in females. However, the EDAC excelled in regards to specificity, producing 96% in males and 90% females (Harasymiw, Vinson, & Bean, 2000).

1.6 Prediction through biomarkers

The objective monitoring of alcohol through the use of biomarkers is opening the door for sophisticated analytics to predict human behavior. The concept of data mining, or the process of discovering unseen patterns or relationships within preexisting data has seen a proliferation with the increase in the amount of data collected as well as advances in computing power (Han, Kamber, & Pei, 2006). Powerful techniques in partnership with new objective techniques of monitoring behavior, is providing tremendous advances in predicting human behavior. The data collected from biomarker interventions, coupled with new patterns in previously collected data are revealing those at a high risk of reoffending. Also, these offenders can be classified and studied in order to ascribe the most appropriate interventions. Moreover, by identifying drivers who might be more likely to reoffend or relapse within their driver's safety plan, clinicians can assign more treatment interventions in an effort to optimize outcomes.

Methods

2.1 Waukesha County Biomarker Pilot

In 2007 and continuing through 2009, The Addiction Resource Council of Waukesha County was the sole site for a pilot study using biomarkers. This study monitored those deemed as high offense drunk drivers for a period of one year. After a driver had committed a third drunk driving offense they were then assessed through this facility and later notified that they would be monitored through biomarkers via the Addiction Resource Council.

During the course of the one year monitoring period, clients were required to report to a local laboratory every three months for a blood sample to be collected. The blood sample would detect the presence of GGT, CDT, as well as EDAC levels, the results would be forwarded to the client's assessor. The assessor would then be able to provide the appropriate interventions based on the levels of these biomarkers and subsequently guide the client's driver's safety plan.

At the time of initial assessment, the client is required to sign a non-disclosure agreement which served to release the results of the their biomarkers to the Addiction Resource Council, and to their physician if medical necessity indicated intervention (Appendix B).

The client completes a routine assessment to assess the severity of alcohol and drug use in the client. Based on the findings of the assessment, additional treatment was recommended in conjunction with being monitored by biomarkers. The combination of treatment as well as biomarker monitoring became the driver's safety plan. In order for the client to successfully complete their driver's safety plan, they would need to adhere to all four biomarker tests as well as adjunctive treatment recommendations. Upon successful completion, the client's driver's license would be reissued. However, if the client did not adhere to their drivers safety plan, the client could have their driver's license revoked indefinitely.

After completing their assessment, the client presented to their local lab facility in order to have their blood drawn, which would be analyzed for the presence of CDT, the assessment of the GGT levels as well as their EDAC result. In addition to these biomarkers, demographic information on the client was collected including, age, gender, days between arrest and assessment (3rd offense), a binary marker if the client committed another re-arrest, employment status, marital status, and timeline follow-back information.

At the time of the initial assessment the Timeline Follow Back (TLFB) was collected as well. The TLFB is an assessment which utilizes a calendar in order to establish the amount of days which an individual consumed alcohol within the past 30 days (Sobell, et. al. 1979). The client is asked to estimate how many

drinks of alcohol that they consumed per day. From this, the TLFB method establishes days drinking, days abstinent, as well as average drinks per day.

As previously stated, the client would present to their local laboratory in order for a blood sample to be obtained for one year. The data would be inputted into a centralized database, where the lab values were recorded as well as the client's demographic information.

2.2 Waukesha Biomarker Dataset

Permission was obtained from the Executive Director of the Addiction Resource Council in order to use the data produced from the pilot (Appendix A). The data was de-identified by the Addiction Resource Council, and given to this researcher via encrypted file. The dataset contained the 249 drivers who participated in this pilot project.

The dataset contained the following variables: days between arrest and assessment, DUI offense number, age, marital status, employment status, initial timeline follow back information, as well as the EDAC, CDT, and GGT results for the client. In 2010, clients were classified as either those who had reoffended (a subsequent DUI in 2011) or no-re-offense based on information obtained by the department of transportation. This binary outcome was coupled with the client's information and was included in this dataset.

In addition to obtaining permission by the Addiction Resource Council, the study protocol was reviewed by the Institutional Review Board at the University of Wisconsin- Milwaukee. The study received a Category 4 exempt status and was approved by IRB# 13.429 (Appendix C).

The data was cleaned and transformed including identification of outliers, as well as identification of missing data. Missing data was coded with a "?" in order to maintain integrity of missing cases.

The data was split into two separate yet similar files in order to evaluate the efficacy of predictions based on how the biomarkers were classified. The biomarker data in file one was kept in continuous form, with the models predicting the target, "re-offense", based on the raw scores from the biomarker results. Biomarker data in file two was transformed into binary values, either the client scored "Positive" or "Negative" based on their biomarker score and the target was again, the binary re-offense or no re-offense variable. The files were both converted into the "ARFF" format. The data was then analyzed using WEKA, an open source software for data mining and machine learning (Hall, Frank, Holmes, Pfahringer, Reutemann, & Witten, 2009).

In the case for file 2, the following cut-offs were used in order to differentiate positive and negative results:

CDT: Greater than 2.2% (Arndt, 2001)

GGT: Greater than 60 units per liter (U/L) (Bianchi , Ivaldi , Raspagni , Arfini , & Vidali , 2010)

EDAC-Test: Greater that 40% (Harasymiw & Bean, 2001)

Objectives

3.1 Objectives for Biomarker Prediction and Clustering

There were two primary objectives within the study, the first was to identify the drinking patterns within the existing biomarker data; and the second, was to predict which individuals were more likely to reoffend, and commit their 4th offense.

1. Identifying drinking patterns

Identify drinking patterns/ treatment patterns within the collected biomarker data. Clients were measured at four distinct points throughout their driver's safety program. At all of these points, CDT, GGT, and EDAC information was collected. These values are distinct in every client. However, it is possible that distinct patterns of drinking were present within groups of clients. For example, a client who was previously a heavy drinker may have abstained upon commencement of his/her driver's safety program. The values also may reflect a

high positive at the initial biomarker collection, and negative values at subsequent tests.

2. Predicting Re-arrest in a high offense population

Utilizing re-arrest data, which will be embedded within the provided dataset by the Addiction Resource Council, analyze predictive factors for the subsequent re-arrest (i.e. Identified drinking pattern, demographics, and biomarker data). This data will highlight a subset of individuals who may be inclined to a further re-arrest. These individuals may have non-compliance within the driver's safety plan, or have continued to consume large amounts of alcohol through the driver's safety plan.

Results

4.1 Group Demographics

The data contained within the file produced an unbalanced dataset, in regards to the binary variable of "re-offense", producing 36 "reoffenders" and 212 " No-Reoffenders." Additionally, the group was 86% male. Of the individuals within the dataset 68% were employed fulltime at assessment, and were 49% single. These groups fall into very distinct clusters, figure 4.1 illustrates the stratification of these clusters. Within the figure, the longer the horizontal bar the

more instances of the combination of the demographic which occurs. For example, single and full time individuals are the most populous combination.

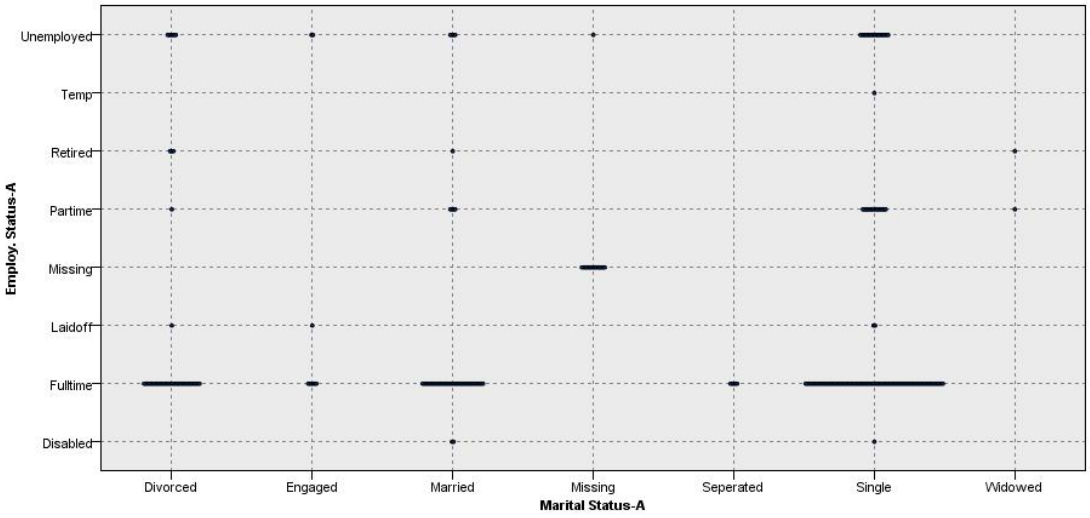


Figure 4.1 Distributions of employment status versus marital statues at assessment

In predicting re-arrest, the target was the binary value, “ Re-offended” or “No Re-offense”. The following inputs were utilized in order predict this value, EDAC values baseline, 3-month,6-month, and final; GGT values baseline, 3-month,6-month, and final; CDT values baseline, 3-month,6-month, and final, days between assess and arrest, timeline follow-back if they self reported abstaining or relapsing, age, martial status, and employment status at the time of arrest.

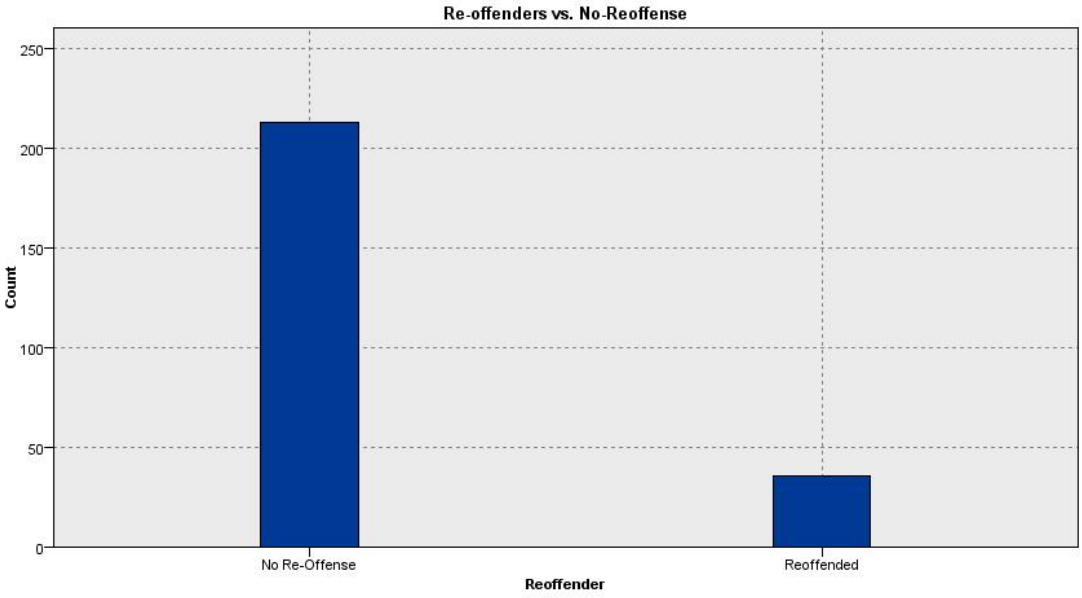


Figure 4.2 Distribution of individuals who reoffended and had no further re-offense

Distribution of age at time of assessment:

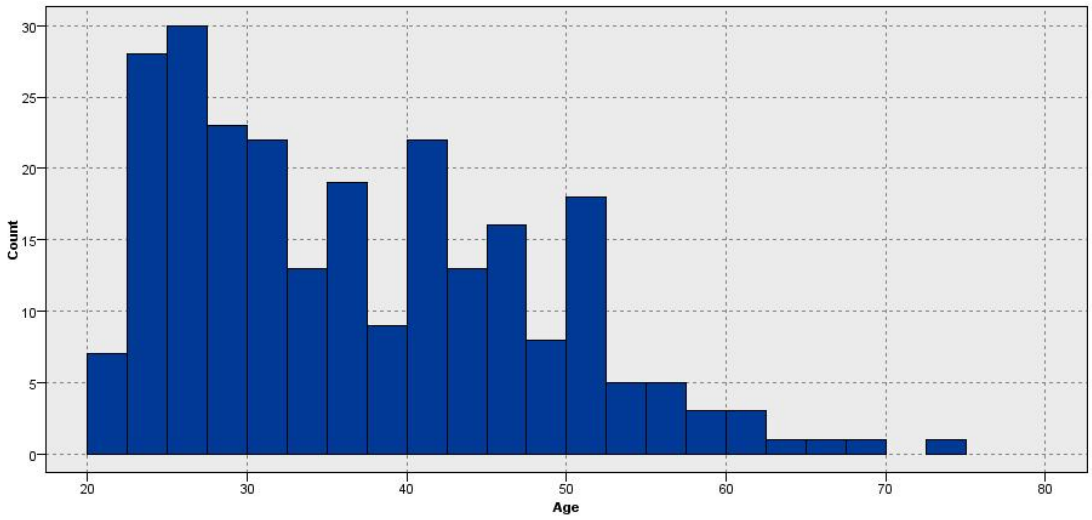


Figure 4.3 Distribution of age at time of assessment

4.2 Predicting Re-arrest

Due to the unbalanced dataset, cost sensitive versions of the classifiers were used which are available in WEKA as the Cost Sensitive Classifier under its meta classifiers. A cost sensitive classifier analyzes the dataset in order to find a predicting scheme that produces the least amount (cost) of errors. A cost matrix

tells the classifier how to weight different types of misclassifications. The matrix below shows the penalizations in a 6:1 ratio used in the experiments which is same as the ratio between “reoffenders” and “no-reoffense”:

	A	B
No-Re-offense	0.0	1.0
Reoffended	6.0	0.0

Table 4.1 Matrix for penalizations of cost sensitive learning classifier

It means that the penalty of a reoffended misclassified as no-re-offense is six times than the penalty of a no-re-offense misclassified as reoffended when the classifier is being trained. Several available base classifiers were tried for the cost sensitive learning classifier, these classifiers were evaluated by how many “re-offenses” the classifier was able to predict based on the dataset. The various classifiers which were tested and their associated predictive power can be found in (Appendix D).

Support Vector Machine classifier was found to produce the highest amount of correct “Re-offense” predictions. Support Vector Machine classifiers, is a supervised learning model, which recognizes patterns within data. The classifier then predicts two possible outcomes based on the associated training data (Cristianini & Shawe-Taylor, 2000). The basic algorithm is based on a non-probabilistic binary linear classification. However, the algorithm can also do non-

linear classification by using non-linear kernels, one of them being the Gaussian Kernel (Press, Teukolsky, Vetterling, & Flannery, 2007) which was found to work best in this study. In this research, the support vector machine classifier analyzes the data and determines if a data point would match either a “Re-offender” or “No Re-Offense”. Within WEKA, the support vector machine classifier utilized Single Minimal Optimization (SMO option in WEKA) numerical method technique. Normalization is important when running these classifiers. The data was automatically normalized before training with the algorithm.

The best results were produced in the file which utilized continuous biomarker data. The classifier was able to accurately predict 64% of the cases based on the aforementioned Support Vector Machine algorithm and using 10-fold cross-validation. In this the data is split into 10 equal parts. Nine parts are used for training and then the trained classifier is tested on the tenth part. This is repeated 10 times with a different test set every time. The results of all the 10 evaluations are then combined and reported. The Support Machine Vector weighted the individual variables; these weights are illustrated in Appendix E.

The confusion matrix for prediction utilizing a cost sensitive learning classifier as well as sequential minimal-optimization algorithm was:

A	B	<-- classified as
96	117	A=No-Re-offense
13	23	B=Re-offended

Table 4.2 Confusion matrix for Cost sensitive learning classifier with SMO base classifier

4.3 Evaluating the accuracy of the prediction

The qualities of the predictions were evaluated based on percentage predicted correct by the confusion matrix. For example, when predicting the re-offended category, the follow equation was utilized based the confusion matrix in figure 4.2:

“B” /total Reffendorsx100= Percent of re-offenders correctly predicted by the cost sensitive classifier.

(23/36) x 100= 64%

In addition to this, the classifier misclassified the No-Reoffense individuals as well. “B”/total No-reoffense x 100.

(117/213) x 100=54%

This researcher also sought to evaluate if more data added will increase the percentage of correctly classified cases. Using the experimental setting within

WEKA this researcher utilized an Instances Results Listener which allows WEKA to vary the amount of training data which each analysis would use. The amount of training data in each analysis was partitioned by percentage : 90, 80, 70, 60, 50, 40, 30, 10. In each partition of the training data, WEKA would utilize the same Support Vector Machine classifier as employed within the previous analysis. Figure 4.4 illustrates the percentages correctly classified by percentage of training data.

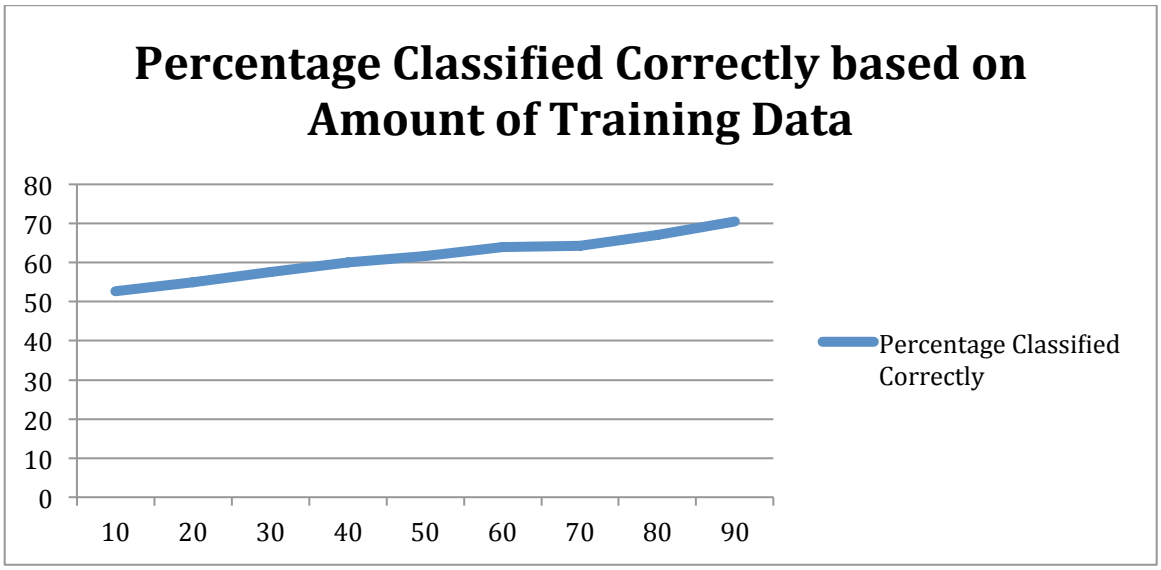


Figure 4.4 Percentage classified correctly

4.4 Value of the predictive input

In addition to this, there were no specific biomarkers which predicted the re-arrest value better then others, with only slightly more importance of the final GGT value. The importance of the biomarker values are visualized below, with the biomarker labeled (EDAC, GGT, or CDT) with the subsequent time period

illustrated (1,2,3, or 4). Figure 4.5 outlines if there were any correlation between the input variables and how important they were to the overall prediction. In the figure, a 1.0 would indicate a great importance to the accuracy of the prediction. Where a 0.0 would indicate no importance of influence to the prediction.

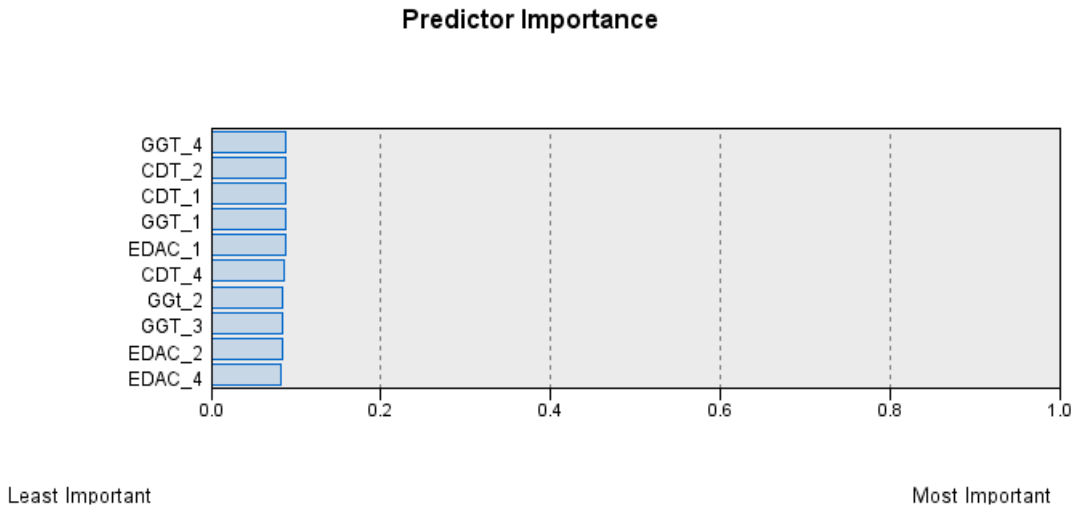


Figure 4.5 Prediction input performance

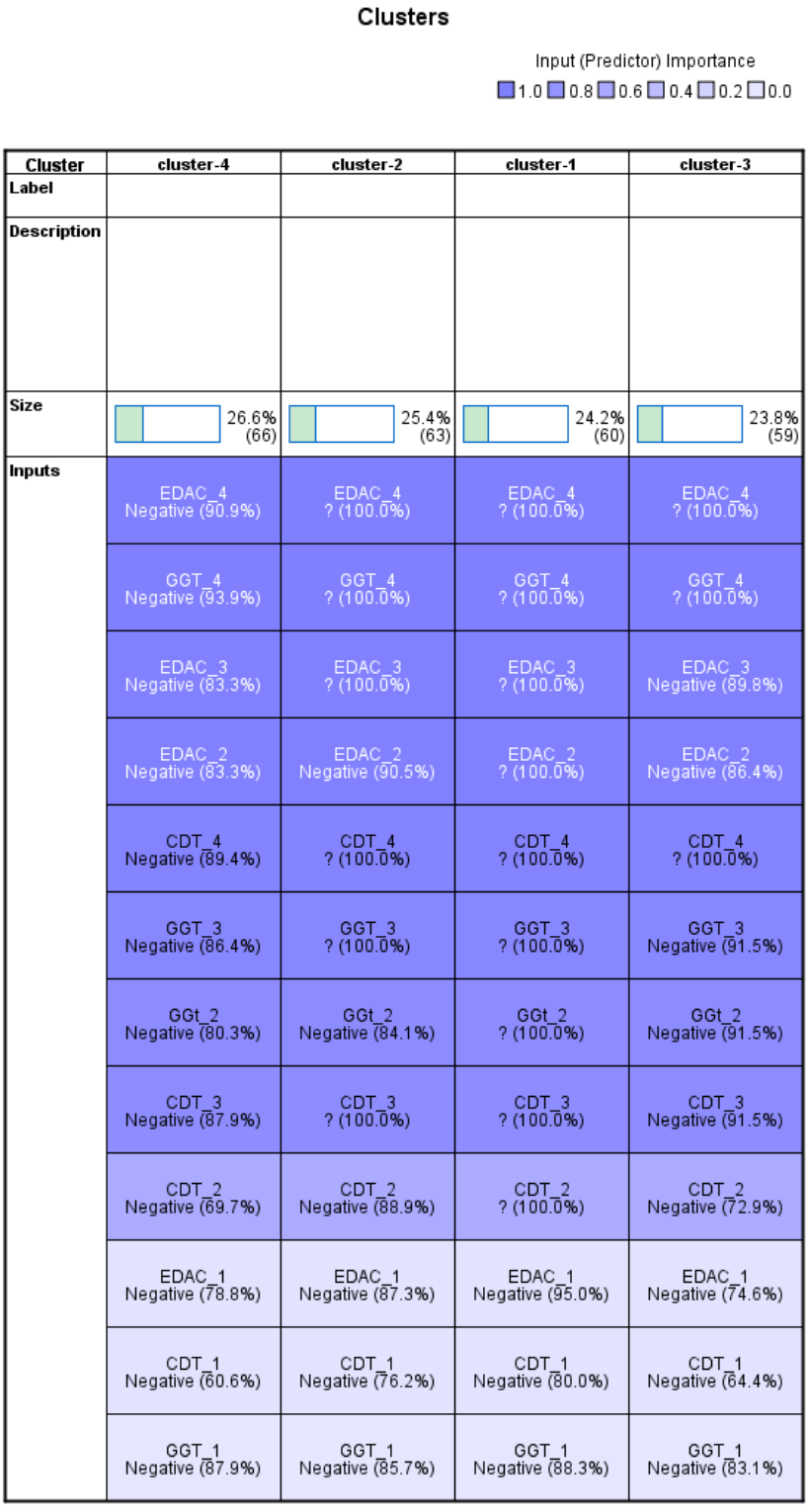
4.5 Clustering individuals throughout the course of monitoring

Biomarker data was processed in order to produce a binary value, " positive" or "Negative" as set by the aforementioned cut off's. This data was then clustered utilizing a TwoStep clustering algorithm (IBM Corporation, 2011). The TwoStep cluster analysis, develops a Cluster Features Tree in order establish baseline nodes. These base lining nodes serve as a summary of the data. After

the tree has been formed, agglomerative clustering is performed in order to produce multiple solutions of the clusters. Agglomerative clustering

This researcher evaluated the cluster based on the silhouette coefficient which illustrates the cohesion of the cluster as well as the separation of the cluster (Kent University). In addition to measuring the clusters as a whole, the silhouette value takes into account the cohesion and separation in the individual data points. The silhouette coefficient value produced was .814 which indicates good separation and tightness of the values.

The cluster assignments are visualized in Figure 4.6. The figure illustrates foremost, the sizes of the clusters. Within the data there are 4 distinct clusters which each roughly make up 25% of the total data. Below is the size and how each input influences the predictor importance of the cluster. In descending order, are the importance of the predictor to the individual clusters with the classification of the predictor, as well as the percentage of individuals who scored that value within the cluster. For example, in cluster 1, 90.9% of individuals had a negative 4th EDAC. This value demonstrates the largest predictor in the formation of the cluster.



Cluster	cluster-4	cluster-2	cluster-1	cluster-3
Label				
Description				
Size	26.6% (66)	25.4% (63)	24.2% (60)	23.8% (59)
Inputs	EDAC_4 Negative (90.9%)	EDAC_4 ? (100.0%)	EDAC_4 ? (100.0%)	EDAC_4 ? (100.0%)
	GGT_4 Negative (93.9%)	GGT_4 ? (100.0%)	GGT_4 ? (100.0%)	GGT_4 ? (100.0%)
	EDAC_3 Negative (83.3%)	EDAC_3 ? (100.0%)	EDAC_3 ? (100.0%)	EDAC_3 Negative (89.8%)
	EDAC_2 Negative (83.3%)	EDAC_2 Negative (90.5%)	EDAC_2 ? (100.0%)	EDAC_2 Negative (86.4%)
	CDT_4 Negative (89.4%)	CDT_4 ? (100.0%)	CDT_4 ? (100.0%)	CDT_4 ? (100.0%)
	GGT_3 Negative (86.4%)	GGT_3 ? (100.0%)	GGT_3 ? (100.0%)	GGT_3 Negative (91.5%)
	GGt_2 Negative (80.3%)	GGt_2 Negative (84.1%)	GGt_2 ? (100.0%)	GGt_2 Negative (91.5%)
	CDT_3 Negative (87.9%)	CDT_3 ? (100.0%)	CDT_3 ? (100.0%)	CDT_3 Negative (91.5%)
	CDT_2 Negative (69.7%)	CDT_2 Negative (88.9%)	CDT_2 ? (100.0%)	CDT_2 Negative (72.9%)
	EDAC_1 Negative (78.8%)	EDAC_1 Negative (87.3%)	EDAC_1 Negative (95.0%)	EDAC_1 Negative (74.6%)
	CDT_1 Negative (60.6%)	CDT_1 Negative (76.2%)	CDT_1 Negative (80.0%)	CDT_1 Negative (64.4%)
	GGT_1 Negative (87.9%)	GGT_1 Negative (85.7%)	GGT_1 Negative (88.3%)	GGT_1 Negative (83.1%)

Figure 4.6 Biomarker cluster analysis

Discussion

5.1 Applications of results

The results of data mining the dataset appear to indicate that there are defined groups of drinkers within the individuals who were monitored by biomarkers. The importance of the inputs indicates that in all clusters the final EDAC was missing. Additionally, there were not enough positive biomarkers which warranted a defined cluster. However, there was enough missing data which warranted a defined cluster.

By having defined groups of drinkers multiple treatment modalities can be established in response to these categories. Drivers who are more inclined to consume alcohol within their drivers safety plan can be allocated more resources or a more intensive drivers safety plan. Having access to this knowledge can also aid in assessors to ascribing the most appropriate treatment as well as evaluating their decisions when terminating an individual's driver's safety plan.

5.2 Limitations of the data

There are also multiple limitations to the study. First, indirect biomarkers may produce inaccurate results. The biomarkers as stated before have multiple limitations based on substances or ailments that may produce false positives or false negatives. Prediction, hinges on the assumptions as well as the validity of

the inputted values, if there are inaccuracies within these values, the accuracy and precision of the prediction may come into question.

However, new alcohol biomarkers are in the process of development, which accurately measure alcohol consumption with extremely low false negatives as well as false positive. These new direct biomarkers are inundating the market and new data is being collected with them as the primary measure. By utilizing these biomarkers in further studies, the inputs can be further verified and subsequently, more accurate predictions can be produced.

Additionally, within the study there was variation between each of the clients testing phases. On average, there was 3 months between when they were scheduled to be tested and when they were actually tested. The variation in this time may not be detrimental, as the test covers a three month drinking history, however this variation may lead to non-accurate predictions. When speaking to members, of the addiction resource council regarding these variations, they stated that this was primarily due to clients missing their appointments due to a variety of reasons. Many times, the assessors I spoke to felt that the clients were delaying the test in order to miss a positive mark. This change in behavior again may vary the results of the predictions. However, assessors felts that if there were more strict guidelines in regards to the programs, such as state level laws and amendments, they would be more inclined to enforce the range of collection times. Again, having less variance between the monitoring periods will lead to better predictions into relapse.

5.3 Limitations of Indirect Biomarkers

Biomarkers are proving to be extraordinary tools in the monitoring of alcohol consumption, however there are limitations inherent in indirect biomarkers. Indirect biomarkers are solely measuring the toxic effects of alcohol on one's system. The biomarkers are not direct measures of alcohol in ones system, and may not be representative of one's true drinking pattern. For example, if an individual has one drinking binge (five drinks or more in a two hour period) in a two-week monitoring period, the biomarker would not show up positive. By failing to detect this drinking pattern, there may be inherent flaws in using indirect biomarkers for predicting rearrests.

5.4 Future Directions

Having information regarding who would relapse, or commit another DUI offense can be invaluable in regards to economic impact, as well as resource allocation. The biomarkers highlighted within this thesis are relatively inexpensive to run. Currently, the EDAC is $36 dollars to perform representing a small value considering the potential return on investment. Clearly, the dataset illustrates that individuals are abstaining or reducing their drinking throughout the monitoring period. This effectiveness is a giant leap in the treatment of these repeat offense drunk drivers.

Furthermore, in the overarching nature of this exercise, this research is attempting to demonstrate the ability to use data mining techniques on complex biomedical data. Much of the data, particularly biomedical data related to behavioral health are analyzed solely with traditional statistical techniques. These techniques are excellent in illustrating apriori hypothesis as well as, limited post-hoc hypothesis. However, within complex data, patterns are inherent which may aid in the evaluation as well as creation of new treatment methods leveraging the power of computers. The research illustrates that we are on the precipitous of this change, and that this new research methods are providing valuable insight within existing biomedical data.

Appendix A

Letter to Use Ex Post Facto/Retrospective Data

5/29/2013

Graduate School
University of Wisconsin-Milwaukee
3203 N Downer Ave
Milwaukee, Wi 53211

University IRB Office:

As Executive Director, I have given Mr. Brian Kay permission to review and use archival data on clients previously enrolled in our biomarker pilot program from 2008-2010. I have spoken with Mr. Kay and understand the scope of his research, and how he will be using our data. All information to be gathered will be done in a confidential, deidentified and in an appropriate manner. Additionally, all data collected will be reported in aggregate under the conditions of the projects Authorization to Disclose Information.

Should you have any questions, please feel free to contact me.

Sincerely,

_____

Dr. Claudia Roska Executive Director of the Addition Resource Council

Appendix B

AUTHORIZATION AND RELEASE FOR ALCOHOL CONSUMPTION TESTING AND MONITORING

I, _____, am a participant in the Addiction Resource Council's ("Agency") Driver Safety Program that monitors my alcohol consumption through the use of bio-markers. By agreeing to participate in this program I hereby agree to the terms and conditions of this Authorization and Release Form ("Form").

1. Consent to Alcohol Consumption Tests and Blood Draws. I hereby agree to undergo alcohol consumption tests ("Testing(s)") at such intervals as established by the Agency to determine my level of alcohol consumption within the 14 to 21 days preceding the Testing. I further agree that the laboratory and Alcohol Detection Services, LLC ("Company") will need accurate information and compliance with the testing procedures from me, in order for them to provide reliable test results:

   (a) I consent to having two vials of my blood drawn for each Testing and authorize the Laboratory the "Laboratory") to run such tests on the blood samples as instructed by the Company for the sole purpose of conducting the Testing(s).

   (b) I authorize the Laboratory to provide the results of my blood test(s) to the Company and to my Primary Care Physician listed below.

   (c) I authorize the Company to conduct the Testing(s) using my blood test results provided by the laboratory and to provide the Testing results directly to the Agency. I understand that Company will not provide me a copy of the testing results and that I must seek information regarding such results directly from the Agency.

   (d) I will be financially responsible for all cost related to each blood draw, Laboratory test and Company testing described in (a-c) above.

   (e) I authorize the Agency, Laboratory, Company and Primary Care Physician to share and communicate with one another as necessary and appropriate for my monitoring and treatment under this program.

2. Primary Care Physician:

Name of Physician: _____

Address: _____

City, State Zip Code: _____

Facsimile Number: _____ Telephone Number: _____

3. List all current Medications you are taking (*Please print*):

_____ _____

_____ _____

_____ _____

_____ _____

4. List all current Medical conditions for which you are being treated (*Please print*):

_____ _____

_____ _____

_____ _____

_____ _____

5. Client current contact information (*Please print*):

Name (Last, First, MI)

_____

Address: _____

Home Telephone Number:

_____

Cell Phone Number:

_____

6. Authorization to Release Confidential Medical Information. I understand that although the Laboratory is subject to state confidentiality laws and the privacy rules under the Health Insurance Portability and Accountability Act of 1996 ("HIPAA"), the Company is not subject to such laws. Whenever possible, Company will comply with the privacy regulations promulgated pursuant to the Health Insurance Portability and Accountability Act of 1996

("HIPAA"). Because the Company is not subject to HIPAA or any state confidentiality laws, I understand that any health information disclosed to the Company pursuant to this Form may be subject to redisclosure and no longer be protected by state confidentiality laws or HIPAA. I further understand that I have the right to revoke this authorization at any time by providing written notice of such revocation to the Agency in accordance with their policies and procedures. I understand that any revocation will not be effective to the extent that any party has already acted in reliance upon this authorization. I authorize and consent for the Company to provide the Testing results to the Agency requesting such Testing(s) or as otherwise required by law. I understand that the Testing results may impact my Driver Safety Plan. This authorization shall be in effect for one year following the date this Form is executed or until I complete my participation in the Agency or complete and am discharged form my Driver Safety Plan, whichever comes first. I also understand that failure to appear at the appointed laboratory to have my blood drawn for the purpose of obtaining and EDAC™ result will be considered a refusal and reported as a positive screen to my attorney and/or the Agency.

7. Release. I understand that the Company is not responsible for any erroneous Testing results that occur because of testing errors made by the Laboratory. I hereby release and forever discharge and hold harmless Company, as well as any of its managers, members, officers, employees, agents and representatives from any claims, liabilities, suits, losses, demands, obligations, costs incurred, expenditures, damages or causes of action of any nature whatsoever arising out of, related to, or in any way connected with the Testing, including without limitation claims, liabilities, suits, losses, demands, obligations, costs incurred, expenditures, damages or causes of action of any nature whatsoever arising from any investigation or personnel actions.

8. General Acknowledgments. By signing below, I acknowledge that I have read this Form and understand the rights I have and the rights I am giving up by agreeing to the terms and conditions set forth in this Form. I also acknowledge that all of the information is true and correct and I have received a copy of this Form.

SIGNATURE: _____ Date: _____

PRINT NAME: _____

AGENCY WITNESS: _____ Date: _____

PRINT NAME: _____

Appendix C

UNIVERSITY of WISCONSIN
UWMILWAUKEE
Department of University Safety & Assurances

Jessica Rice
IRB Administrator
Institutional Review Board
Engelmann 270
P. O. Box 413
Milwaukee, WI 53201-0413
(414) 229-3182 *phone*
(414) 229-6729 *fax*

http://www.irb.uwm.edu
ricej@uwm.edu

New Study - Notice of IRB Exempt Status

Date: June 17, 2013

To: Rohit Kate, PhD
Dept: College of Health Sciences

Cc: Brian Kay

IRB#: 13.429
Title: ALCOHOL BIOMARKERS AS PREDICTIVE FACTORS OF REARREST IN HIGH RISK REPEAT OFFENSE DRUNK DRIVERS

After review of your research protocol by the University of Wisconsin – Milwaukee Institutional Review Board, your protocol has been granted Exempt Status under Category 4 as governed by 45 CFR 46.101(b).

Unless specifically where the change is necessary to eliminate apparent immediate hazards to the subjects, any proposed changes to the protocol must be reviewed by the IRB before implementation. It is the principal investigator's responsibility to adhere to the policies and guidelines set forth by the UWM IRB and maintain proper documentation of its records and promptly report to the IRB any adverse events which require reporting.

It is the principal investigator's responsibility to adhere to UWM and UW System Policies, and any applicable state and federal laws governing activities the principal investigator may seek to employ (e.g., FERPA, Radiation Safety, UWM Data Security, UW System policy on Prizes, Awards and Gifts, state gambling laws, etc.) which are independent of IRB review/approval.

Contact the IRB office if you have any further questions. Thank you for your cooperation and best wishes for a successful project

Respectfully,

Jessica P. Rice
IRB Administrator

Appendix D

Classifier	Percent "Re-offender" Classified Correctly	Percent "No Re-Offense Classified Correctly	Number of folds
SVM	64%	55%	10
Neural Network	17%	19%	10
Logistic Regression	25%	36%	10
Decision Tree (J 48)	0%	0%	10
Classify and Regression Tree	0%	0%	10

Appendix E

```
      0.8955 * (normalized) Days btw
+     -2.0574 * (normalized) DUI
+      0.1849 * (normalized) EDACS1
+     -0.3676 * (normalized) EDACT1
+      0.6553 * (normalized) GGT1
+     -0.2897 * (normalized) CDT1
+      0.1139 * (normalized) EDACS2
+      0.2697 * (normalized) EDACT2
+      0.2484 * (normalized) GGT2
+     -1.5089 * (normalized) CDT2
+      0.4252 * (normalized) EDACS3
+      0.3726 * (normalized) EDACT3
+     -0.0899 * (normalized) GGT3
+     -0.0966 * (normalized) CDT3
+      0.1686 * (normalized) EDACS4
+      1.1438 * (normalized) EDACT4
+      0.6142 * (normalized) GGT4
+      0.4973 * (normalized) CDT4
+      1.5009 * (normalized) TLFB
+      0.0156 * (normalized) Total drinks
+      0.0699 * (normalized) mean drinks
+     -0.8163 * (normalized) drinking days
+      0.176  * (normalized) abstinent days
+     -1.049  * (normalized) Gender
+     -1.2289 * (normalized) Age
+      0.608  * (normalized) Employment=Unemployed
+     -0.0462 * (normalized) Employment=Part time
+      0.0836 * (normalized) Employment=Fulltime
+      1.0958 * (normalized) Employment=Retired
+     -0.5804 * (normalized) Employment=Laid off
+     -0.5804 * (normalized) Employment=Temp
+     -0.5804 * (normalized) Employment=Disabled
+      0.1079 * (normalized) Marital=Single
+     -0.2003 * (normalized) Marital=Married
+      1.6348 * (normalized) Marital=Engaged
+     -0.5804 * (normalized) Marital=Widowed
+      0.396  * (normalized) Marital=Divorced
+     -1.3578 * (normalized) Marital=Separated
-     1.387
```

Works Cited

Appenzeller, B. M., Schneider, S., Maul, A., & Wennig, R. (2005, August). Relationship between blood alcohol concentration and carbohydrate-deficient transferrin among drivers. *Drug and Alcohol Dependence* , 261-265.

Arndt, T. (2001). Carbohydrate-deficient Transferrin as a Marker of Chronic Alcohol Abuse: A Critical Review of Preanalysis, Analysis, and Interpretation. *Clinical Chemistry , 47* (1), 13-27.

Bean, P., Roska, C., Harasymiw, J., Pearson, J., Kay, B., & Louks, H. (2009). Alcohol Biomarkers as Tools to Guide and Support Decisions About Intoxicated Driver Risk. *Traffic Injury Prevention , 10* (6), 519-527.

Bianchi , V., Ivaldi , A., Raspagni , A., Arfini , C., & Vidali , M. (2010). Use of carbohydrate-deficient transferrin (CDT) and a combination of GGT and CDT (GGT-CDT) to assess heavy alcohol consumption in traffic medicine. *Alcohol & Alcohol Research , 45* (3), 247-251.

Couture, S., Brown, T. G., Tremblay, J., Kin, N., Ouimet, M., & Nadeau, L. (2012). Are biomarkers of chronic alcohol misuse useful in the assessment of DWI recidivism status? *Accident Analysis and Prevention , 42* (1), 307-312.

Cristianini, N., & Shawe-Taylor, J. (2000). *AN Introduction to Support Vector Machines and other kernel-based learning methods.* Cambridge University Press.

Elder, R., Voas, R., Beirness, D., Shults, R. A., Sleet, D. A., Nichols, J. L., et al. (2011). Effectiveness of Ignition Interlocks for Preventing Alcohol-Impaired Driving

and Alcohol-Related Crashes : A Community Guide Systematic Review. *American Journal of Preventive Medicine , 40* (3), 362-376.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. (2009). The WEKA Data Mining Software: An Update. *SIGKDD Explorations , 11* (1).

Han, J., Kamber, M., & Pei, J. (2006). *Data Mining: Concepts and Techniques.* Waltham, MA: Elsevier.

Harasymiw, J., & Bean, P. (2001). The combined use of the early detection of alcohol consumption (EDAC) test and carbohydrate-deficient transferrin to identify heavy drinking behaviour in males. *Alcohol ad Alcoholism , 36* (4), 349-353.

Harasymiw, J., Vinson, D. C., & Bean, P. (2000). The early detection of alcohol consumption (EDAC) score in the identification of Heavy and at-risk drinkers from routine blood tests. *Journal of Addiction and Diseases , 19* (3), 43-59.

IBM Corporation. (2011). *IBM SPSS Statistics* . Retrieved April 24, 2013, from IBM SPSS Statistics : http://publib.boulder.ibm.com/

Javors, M., & Johnson, B. A. (2003). Current status of carbohydrate-deficient transferrin, total serum sialic acid, sialic acid index of apolipoprotein J and serum beta-hexosaminidase as markers for alcohol consumption. *Addiction , 98* (2), 45-50.

Kent University . (n.d.). *Cluster Validation.* Retrieved June 1, 2013, from Kent.edu: http://www.cs.kent.edu/~jin/DM08/ClusterValidation.pdf

Korzec, A., Bär, M., Koeter, M. W., & Kieviet, W. (2001). DIAGNOSING ALCOHOLISM IN HIGH-RISK DRINKING DRIVERS: COMPARING DIFFERENT DIAGNOSTIC

PROCEDURES WITH ESTIMATED PREVALENCE OF HAZARDOUS ALCOHOL USE. *Alcohol and Alcoholism , 36* (6), 594-602.

Marques, P., Tippetts , S., Allen , J., Javors , M., Alling , C., Yegles , M., et al. (2009). Estimating driver risk using alcohol biomarkers, interlock blood alcohol concentration tests and psychometric assessments: Initial descriptives. *Addiction Methods and Techniques , 10* (11), 226-239.

Peterson , K. (2004/2005). Biomarkers for Alcohol Use and Abuse. *Alcohol Research & Health , 28* (1), 30-37.

Press, W., Teukolsky, S., Vetterling, W., & Flannery, B. (2007). *Numerical Recipes: The Art of Scientific Computing (3rd ed.).* New York, NY: Cambridge University Press.

Rossman, A. S., & Lieber, C. S. (1990). Biochemical markers of alcohol consumption. *Alcohol Research , 14*, 210-218.

Sobell, L. C., Sobell, M. B., & VanderSpek , R. (1979). Timeline Followback (TLFB) for AlcoholRelationships between clinical judgment, self- report and breath analysis measures of intoxication in alcoholics. *Journal of Consulting and Clinical Psychology , 47*, 204-206.

Voas, R. B., Tippets , S. S., Fisher, D., & Grosz, M. (2010). Requiring suspended drunk drivers to install alcohol interlocks to reinstate their licenses: effective? *Addiction , 105* (8), 1422-1427.

Walker, L. (2012, June 19). Waukesha's alcohol biomarker experiment shows positive results. *Journal Sentinal* .

Wisconsin Department of Transportation. (2010, August). *Frequently asked questions about ignition interlock devices .* Retrieved March 1, 2013, from Wisconsin Department of Transportation: http://www.dot.wisconsin.gov/statepatrol/docs/iid-faq.pdf

CURRICULUM VITAE

Brian Kay

Place of birth: Libertyville, IL

Education:

B.A., Marquette University, May 2009

Major: Psychology

Thesis Title: Alcohol Biomarkers as Predictive Factors of Rearrest in High Risk Repeat Offense Drunk Drivers

Publications:

- Long-term effects of a multidisciplinary residential treatment model on improvements of symptoms and weight in adolescents with eating disorder. *Journal of Groups in Addiction & Recovery*. *Article in review.*
- Recidivism Risk of Repeat Intoxicated Drivers Monitored with Alcohol Biomarkers. *Traffic Injury Prevention*. *Article in review.*
- Clinical Observation of the Impact of Maudsley therapy in Improving Eating Disorder Symptoms, Weight, and Depression in Adolescents Receiving Treatment for Anorexia Nervosa. *Journal of Groups in Addiction & Recovery*, Volume 5, Issue 1, 70.
- Alcohol Biomarkers as Tools to Guide and Support Decisions About Intoxicated Driver Risk. *Traffic Injury Prevention*, Volume 10, Issue 6, 519

Invited Poster Presentations:

- A Pilot Study of Cognitive Behavioral Therapy as a Treatment Adjunct for Eating Disordered Patients with Co-Morbid Anxiety: A Comparison with Treatment-As-Usual. Poster presentation at The International Conference on Eating Disorders, Austin, TX, May 3, 2012.
- Comparison of Adults and Adolescents Patients? Profiles at Admission to Residential Eating Disorder Treatment. Poster presentation at The International Conference on Eating Disorders, Austin, TX, May 3, 2012.
- Symptoms Severity, Demographics and Weight Profile of Anorexic Patients Admitted to Eating Disorder Treatment Across the Continuum of Care. Poster presentation at The International Conference on Eating Disorders 2009.