Paper SA-11-2012

ROC Analysis for the Evaluation of Ethyl Glucuronide (EtG) as a Long-term Alcohol Biomarker in Hair and Fingernails: Evidence from a Large College Drinking Study

Daniel Fuhrmann[1], Michael Fendrich[1], Lisa Berger[1]
Charles Plate[2], Doug Lewis[2], Joseph Jones[2]
[1]Center for Applied Behavioral Health Research, University of Wisconsin-Milwaukee, Milwaukee, WI
[2]United States Drug Testing Laboratories, Inc., Des Plaines, IL

ABSTRACT

Previous studies on human hair samples have shown mixed results regarding the diagnostic accuracy of the biomarker Ethyl glucuronide (EtG) in screening subjects for heavy alcohol use. In contrast, the use of human fingernails to detect EtG has only been explored in one study and the utility of fingernail EtG tests as a screening tool for individuals with problematic levels of alcohol consumption is currently unknown.

This paper seeks to narrow that gap in the literature by contrasting the continuous biomarkers hair and fingernail EtG with a modified diagnostic criterion of "presence of alcohol abuse or dependence" that was constructed using the Mini International Neuropsychiatric Interview (M.I.N.I.) and DSM-IV criteria. The data for the present study were collected from a random sample of 606 undergraduate students from a large, public urban university. The paper reports early analysis results from what is still an ongoing research process and focuses on the statistical approaches that were employed using various SAS/STAT® procedures and SAS® macros.

KEYWORDS

Sensitivity, specificity, diagnostic accuracy, receiver operating characteristic curves, ROC, diagnostic threshold, biomarkers, FREQ procedure, LOGISTIC procedure

INTRODUCTION

Ethyl glucuronide (EtG) is a metabolite of ethanol that can be detected in human hair as it is deposited into the hair shaft through sweat. Previous analyses of the accuracy of EtG in human hair to screen for heavy alcohol use have shown positive, yet mixed, results. Many studies based on samples of drinkers from different populations have demonstrated the successful use of EtG in hair to detect chronic heavy drinking (Pragst et al. 2010, among others). A recent study of 100 people with a range of alcohol use, however, came to the conclusion that EtG in hair was a low sensitivity marker for alcohol exposure and should not be used quantitatively to indicate alcohol consumption (Lees et al., 2012). In contrast, the uptake of EtG in human fingernails to detect EtG has only been discussed in one study (Jones et al., 2012) and the utility of fingernail EtG tests as a detection tool for individuals with dangerous levels of alcohol consumption has not been analyzed yet.

The data for the present study were obtained from a random sample of N=606 undergraduate students from a large, public urban university. Survey instruments such as the calendar-based Timeline Follow Back method were used to collect self-reported data on alcohol drinking patterns and consumption levels during the 90-day time period preceding the interview day. Based on responses to the structured diagnostic interview M.I.N.I. (Sheehan et al., 2009), study participants were classified for the presence (or absence) of alcohol abuse or dependence based on DSM-IV criteria (American Psychiatric Association, 2000). Hair and fingernail specimens were collected from the majority of study participants and tested for their EtG concentration (see Jones et al., 2012, for more detail). The original sample (N=606) was reduced by one observation per list-wise deletion as the corresponding EtG hair result was identified as a univariate outlier. Furthermore, the present analyses are based on a sample of matched pairs of hair and fingernail EtG concentrations (n=530).

This paper reports early analysis results from what is still an ongoing research process and it focuses on the statistical approaches that were employed using SAS ® software. In order to not detract from this objective, a simplified outcome of "presence of alcohol abuse or dependence" has been constructed to demonstrate the analytical tools that Base SAS® 9.3 and SAS/STAT® 9.3 offer in the context of biomarker research.[1]

[1] Note that the combination of these two outcomes (alcohol abuse and dependence) into one does not follow DSM-IV protocol.

In analyzing the two biomarkers hair and fingernail EtG, the discriminating power of each detection tool was first assessed based on sensitivity, specificity, positive and negative predictive value at pre-defined threshold values using the FREQ procedure. Receiver operating characteristic (ROC) curves as part of the LOGISTIC procedure were subsequently used to compare the two biomarkers against each other as they were associated with the outcome. Lastly, optimal thresholds for classification based on test results for each biomarker were determined using the LOGISTIC procedure in combination with the SAS® macro *%rocplot*. Partial results are explained in their respective section, and a summary of the findings and study limitations can be found in the discussion. A description of how we plan to expand on the present findings marks the conclusion of the paper.

BASIC MEASURES OF DIAGNOSTIC ACCURACY

There exists a large variety of statistical methods to assess the accuracy of diagnostic tests in depicting the presence of a particular disease or condition. Their applicability largely depends on whether the data are continuous (Pepe, 2003; Krzanowski & Hand, 2009) or categorical (Zou et al., 2012; Zhou et al., 2012) but the two most common measures of diagnostic accuracy that can be applied to both kinds of data are sensitivity and specificity. Both are best illustrated by using a decision matrix as illustrated in Figure 1.

Figure 1: Decision matrix

Test Result	Disease / Condition		
	Absent	Present	Total
Negative	True Negative (TN)	False Negative (FN)	TN+FN
Positive	False Positive (FP)	True Positive (TP)	TP+FP
Total	TN+FP	TP+FN	TP+FP+TN+FN=N

Sensitivity is the fraction of true positives divided by the sum of true positives and false negatives: $Se = \frac{TP}{TP+FN}$. It is also referred to as the true-positive fraction (TPF) and is interpreted as the conditional probability of a positive test given that the subject has the disease or condition.[2] The false-positive fraction (FPF) is 1-TPF.

Specificity on the other hand is the ratio of true negatives divided by the sum of true negatives and false positives: $Sp = \frac{TN}{TN+FP}$. Another term for specificity is the true-negative fraction (TNF) which represents the conditional probability that a negative test is obtained given that the subject does not have the disease or condition. The false-negative fraction (FNF) is 1-TNF.

In the context of prediction, two additional measures related to the decision matrix above come to mind. The first is the proportion of subjects for whom the test result is positive but who do not actually have the disease or condition. It is the ratio of true positives divided by the sum of true positives and false positives and is called the positive predictive value: $PPV = \frac{TP}{TP+FP}$. The corresponding metric that expresses the ratio of true negatives divided by the sum of true and false negatives is the negative predictive value: $NPV = \frac{TN}{TN+FN}$.
A measure that summarizes the overall misclassified observations is derived by dividing the sum of false positives and false negatives by the overall sample size. It is referred to as the misclassification rate: $MR = \frac{FP+FN}{TN+FP+FN+TP}$.

In the context of medical diagnostics, a test with poor predictive capability of the actual disease status or condition of a patient can lead to negative – and potentially costly – consequences. For example, a healthy person could be falsely diagnosed with having an infectious disease if the test lacked specificity, and as a consequence, the individual might incur both emotional and financial stress. On the contrary, a diseased individual might be tested negative (i.e. not having the disease) due to low test sensitivity, and might end up infecting other people with the disease without his/her knowledge. Ultimately, wrongful classification of a person may even have serious legal consequences and developers of diagnostic tests need to take these tradeoffs into consideration.

DIAGNOSTIC ACCURACY AT PREDEFINED THRESHOLDS

In this section, we demonstrate a common analysis of accuracy using two binary variables. Deriving the measures described above is a straight-forward task in SAS® when both the marker and the outcome are dichotomous. Using our study data, frequencies of the alcohol abuse/dependence variable as well as the hair and fingernail EtG biomarker variables are displayed in Table 1. Both biomarkers were measured on a continuous scale but the hair EtG values were dichotomized using a cut-off value of 30 (pg/mg) as proposed by the Society of Hair Testing (SOHT, 2011), while the fingernail EtG cut-off value was set at 20 (pg/mg). Out of the 530 students in the analysis sample, 205 (38.7%) satisfied the criterion for either alcohol abuse or dependence. Only 46 students had hair EtG values above the predefined threshold value of 30 (pg/mg), but 178 students exceeded the cut-off value of 20 (pg/mg) for the fingernail test.

[2] The terms “fraction” and “rate” are being used interchangeably in the literature in the context of diagnostic accuracy. For example, it is common to refer to the TPF as the true-positive rate (TPR).

Table 1: Frequencies of analysis variables

Frequencies of Abuse/Dependence Analysis Variables

The FREQ Procedure

Alcohol Abuse or Dependence

diag	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No Abuse or Dependence	325	61.32	325	61.32
Abuse or Dependence	205	38.68	530	100.00

Hair EtG

hair_etg	Frequency	Percent	Cumulative Frequency	Cumulative Percent
< 30 pg/mg	484	91.32	484	91.32
>= 30 pg/mg	46	8.68	530	100.00

Fingernail EtG

nail_etg	Frequency	Percent	Cumulative Frequency	Cumulative Percent
< 20 pg/mg	352	66.42	352	66.42
>= 20 pg/mg	178	33.58	530	100.00

The decision matrix from Figure 1 can be derived using the FREQ procedure. We are showing the code to create the 2*2 contingency table for one of the biomarkers only but are displaying the output for the second marker as well in Table 2 below.

```
PROC FORMAT;
    VALUE hair_fmt      0 -< 30 = '< 30 pg/mg'
                        30 - high = '>= 30 pg/mg';

        VALUE nail_fmt  0 -< 20 = '< 20 pg/mg'
                        20 - high = '>= 20 pg/mg';

        VALUE diag_fmt  0 = 'No Abuse or Dependence'
                        1 = 'Abuse or Dependence';

    VALUE diag_ynfmt    0 = 'No'
                        1 = 'Yes';
RUN;

PROC FREQ DATA=etg_matched;
TABLES hair_etg*diag / NOPERCENT;
FORMAT diag diag_fmt. hair_etg hair_fmt.;
LABEL diag='Alcohol Abuse or Dependence' hair_etg='Hair EtG';
RUN;
```

Table 2: Decision matrices for hair or fingernail EtG results by diagnosis of alcohol abuse/dependence

Table of hair_etg by diag

hair_etg(Hair EtG) diag(Alcohol Abuse or Dependence)

Frequency / Row Pct / Col Pct	No Abuse or Dependence	Abuse or Dependence	Total
< 30 pg/mg	307 63.43 94.46	177 36.57 86.34	484
>= 30 pg/mg	18 39.13 5.54	28 60.87 13.66	46
Total	325	205	530

Table of nail_etg by diag

nail_etg(Fingernail EtG) diag(Alcohol Abuse or Dependence)

Frequency / Row Pct / Col Pct	No Abuse or Dependence	Abuse or Dependence	Total
< 20 pg/mg	258 73.30 79.38	94 26.70 45.85	352
>= 20 pg/mg	67 37.64 20.62	111 62.36 54.15	178
Total	325	205	530

The numbers in the top row of each cell inside of the tables represent frequencies, the second numbers are row percentages and the last rows display the column percentages. The NOPERCENT option in the TABLES statement suppressed the cell percentages.

For the hair test, the sensitivity of the biomarker is very low ($Se = \frac{28}{205} = 13.66\%$) while the specificity is high ($Sp = \frac{307}{325} = 94.46\%$). That means that a person who is positive for alcohol abuse or dependence would only be detected by

the hair test with a probability of 13.66% when the predefined threshold is used as a cut-off. A person without one of the diagnoses, however, has a probability of 94.46% to be correctly classified by the test.
The PPV of the hair test is 60.87% (28/46) and the NPV is 63.43% (307/484).
The fingernail test appears to perform better than the hair test in terms of sensitivity ($Se = \frac{111}{205} = 54.15\%$), but its specificity is lower ($Se = \frac{258}{325} = 79.38\%$). The point estimate of PPV=62.36% (111/178) of the fingernail test is comparable to that of hair but its NPV is higher at 73.3% (258/352).
The overall misclassification rates of hair and fingernail tests are $\frac{177+18}{530} = 36.79\%$ and $\frac{94+67}{530} = 30.38\%$, respectively.

In order to obtain the corresponding asymptotic standard errors (ASE) for our sensitivity and specificity estimates, we need to subset our data in another run of PROC FREQ. The confidence intervals based on asymptotic theory as well as exact binomial confidence limits can be obtained with the BINOMIAL option in the TABLES statement paired with the EXACT BINOMIAL statement:

```
PROC SORT DATA=etg_matched;
BY DESCENDING hair_etg;
RUN;

TITLE 'Sensitivity of Hair EtG Test for Abuse/Dependence';
PROC FREQ DATA=etg_matched(where=(diag=1)) ORDER=data;
TABLES hair_etg;
EXACT BINOMIAL;
FORMAT diag diag_fmt. hair_etg hair_fmt.;
LABEL diag='Alcohol Abuse or Dependence' hair_etg='Hair EtG';
RUN;
```

The code above sorts the sample data set in descending order by the hair test values first. The FREQ procedure than generates the confidence limits. The WHERE= data set option is used to subset the data to only those observations that fell into the positive group for the diagnoses of alcohol abuse or dependence. The ORDER=data option ensures that the confidence limits are calculated for the proportion of observations that tested positive which appear first in the sorted data set. By default, the confidence limits are calculated for the proportions contained in the first row of the contingency table. The code for the fingernail test is omitted here but the output is presented below.

Table 3: Sensitivity of hair and fingernail tests with confidence limits

Sensitivity of HAIR EtG Test for Abuse/Dependence

The FREQ Procedure

Hair EtG

hair_etg	Frequency	Percent	Cumulative Frequency	Cumulative Percent
>= 30 pg/mg	28	13.66	28	13.66
< 30 pg/mg	177	86.34	205	100.00

Binomial Proportion for hair_etg = >= 30 pg/mg

Proportion (P)	0.1366
ASE	0.0240
95% Lower Conf Limit	0.0896
95% Upper Conf Limit	0.1836
Exact Conf Limits	
95% Lower Conf Limit	0.0927
95% Upper Conf Limit	0.1913

Test of H0: Proportion = 0.5

ASE under H0	0.0349
Z	-10.4066
One-sided Pr < Z	<.0001
Two-sided Pr > |Z|	<.0001
Exact Test	
One-sided Pr <= P	5.866E-28
Two-sided = 2 * One-sided	1.173E-27

Sample Size = 205

Sensitivity of Fingernail EtG Test for Abuse/Dependence

The FREQ Procedure

Fingernail EtG

nail_etg	Frequency	Percent	Cumulative Frequency	Cumulative Percent
>= 20 pg/mg	111	54.15	111	54.15
< 20 pg/mg	94	45.85	205	100.00

Binomial Proportion for nail_etg = >= 20 pg/mg

Proportion (P)	0.5415
ASE	0.0348
95% Lower Conf Limit	0.4733
95% Upper Conf Limit	0.6097
Exact Conf Limits	
95% Lower Conf Limit	0.4706
95% Upper Conf Limit	0.6111

Test of H0: Proportion = 0.5

ASE under H0	0.0349
Z	1.1873
One-sided Pr > Z	0.1175
Two-sided Pr > |Z|	0.2351
Exact Test	
One-sided Pr >= P	0.1319
Two-sided = 2 * One-sided	0.2637

Sample Size = 205

The confidence limits for hair and fingernail sensitivity do not overlap, which indicates a significant difference between the test sensitivities. Application of McNemar's test of equality of paired proportions is not presented here.

Specificity estimates and their confidence intervals for both biomarkers were obtained using similar code. The only differences were that the data had to be subset to include only those students who did not have alcohol abuse or dependence and that the ORDER=freq option was used to place specificity in the first row of the PROC FREQ output (omitted here). The 95% confidence intervals based on exact binomial distribution theory were [0.914, 0.967] for hair and [0.746, 0.837] for fingernails, suggesting that hair tests are more specific than fingernail tests.

The question arises, however, what the appropriate techniques of determining diagnostic accuracy of a marker are when no predetermined cut-off value is known. As Obuchowsky et al. (2004) and Shin (2009) point out, the approach described in this paper so far is only valid when a threshold value to dichotomize the marker had been predefined. While it is not the scope of this paper to compare both biomarkers in terms of their sensitivity at predefined specificity and vice versa, it is of interest to explore whether fingernail testing is more accurate in depicting alcohol-related outcomes than hair testing. Instead of dichotomizing the biomarker variables, the full range of the continuous biomarker values is used in the following exploration of ROC curves. All analyses are based on the joint outcome of "presence of alcohol abuse or dependence."

INDIVIDUAL ROC CURVES AND COMPARISON OF TWO CORRELATED ROC CURVES

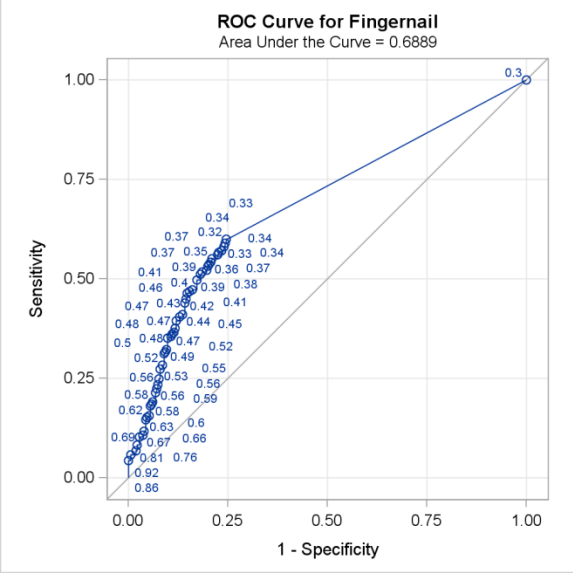
Instead of predicting a binary outcome with a binary variable, we are now focusing on a continuous predictor. Using the full range of a continuous predictor is far more efficient than dichotomization because no information is lost (Gönen, 2006 & 2007). Consider moving the threshold for the continuous biomarker from lowest to highest value in the data, leading to a number of sensitivity-specificity pairs equal to the number of unique biomarker values. The empirical receiver operating characteristic curve is then a two-dimensional plot of all pairs of TPF (sensitivity) on the vertical axis and FPF (1-specificity) on the horizontal axis across the whole range of the continuous predictor values. Since both sensitivity and specificity range from 0 to 1 (or 0 to 100%), the ROC curve is displayed in a quadrant of side length equal to 1. A point of interest in this quadrant is (0, 1) in the top-left corner as it indicates the point of theoretically perfect accuracy (sensitivity=specificity=1). Another useful addition to the graph is a 45-degree line that represents the ROC curve for a test that is no better than chance (equivalent to a random coin toss).

PROC LOGISITIC is used to create the ROC analysis. PLOTS=ROC instructs ODS GRAPHICS to only plot the ROC curves and to omit other graphical output related to the LOGISTIC procedure. The LOGISTIC procedure generates the values for the empirical ROC curve when the ROC statement is invoked. Since parameter estimates and model goodness-of-fit statistics will be generated by the ROC statement, the NOFIT option in the MODEL statement is used to force SAS to ignore the model listed. The code below generates ROC curves for each predictor individually (as two ROC statements are used) and runs the comparison of correlated ROC curves based on the nonparametric approach proposed by DeLong et al. (1988) via the ROCCONTRAST statement. Note that each ROC analysis has been assigned a name in quotation marks immediately following the ROC statement and the respective analysis variable comes last. A reference ROC curve has to be specified for the contrast by using the REFERENCE option and the name of the curve in parentheses. The ESTIMATE option in the ROCCONTRAST statement computes and tests each ROC comparison and the E option displays the contrast coefficients.

```
ODS GRAPHICS ON;
PROC LOGISTIC DATA=etg_matched PLOTS=ROC(ID=prob);
MODEL diag(EVENT='1') = hair_etg nail_etg / NOFIT;
ROC 'Hair' hair_etg;
ROC 'Fingernail' nail_etg;
ROCCONTRAST REFERENCE('Hair') / ESTIMATE E;
RUN;
ODS GRAPHICS OFF;
```

Partial output from the above code is displayed next. The individual ROC analysis for hair tests is omitted here.

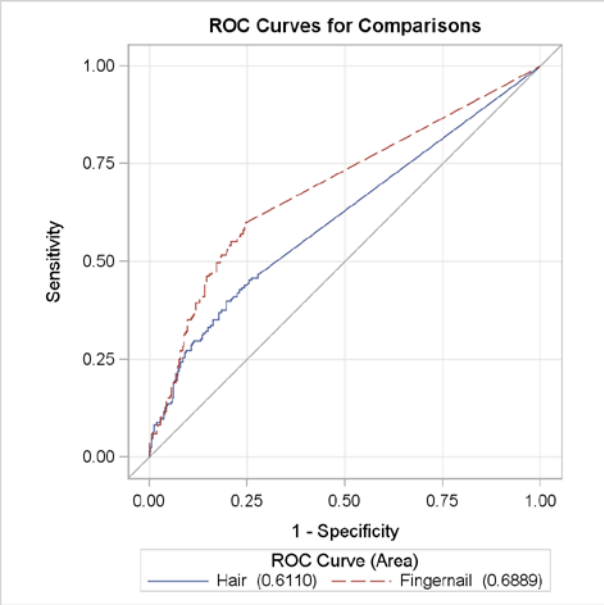
Figure 2: ROC curve for fingernail test



Note that the ROC plot produced by ODS GRAPHICS labels the points on the ROC curve with the predicted probabilities as requested by the ID= option. But while it is important to assess the predicted probability of a random draw from the sampling distribution and to assess its probability of being in the higher class (presence of alcohol abuse or dependence), it is often convenient to refer to a summary measure of the ROC curve. One of the key measures of overall performance of a biomarker is the area under the curve (AUC). The AUC represents the average sensitivity of the biomarker over the whole range of specificities and is the same as the c statistic (concordance) in PROC LOGISTIC. In mathematical terms, AUC is simply the surface area enclosed by the ROC curve and the 45-degree line which can be computed using integration. As seen in Figure 2, the AUC for fingernail tests is 0.69, which is a fairly high value for a biomarker, indicating that the test will be better than chance alone in determining presence of alcohol abuse or dependence.

However, how would we answer the question whether fingernail EtG tests are globally a better biomarker for heavy alcohol use than hair tests? Graphically, we are able to deduce from Figure 3 that the ROC curve for fingernails lies on top or above the ROC curve for hair across all levels of FPF. In other words, the fingernail test is either equally or more sensitive than the hair test at any given level of specificity. This is further supported by the lower AUC for hair (0.61) than for fingernails (0.69). But is the difference statistically significant?

Figure 3: Comparison of correlated ROC curves for hair and fingernail biomarkers



The tables summarized in Table 4 are the result of the ROCCONTRAST statements. The table ROC Association Statistics provides both biomarker's AUCs including their standard errors and 95% confidence limits to account for sampling variability. The table ROC Contrast Coefficients was produced by the E option and displays Hair as the reference group (coded as -1). The corresponding table ROC Contrast Test Results indicates that the 1-df Chi-square test of equal AUCs rejects that null hypothesis at the 0.01 significance level.

Table 4: Output from analysis of correlated ROC curves

ROC Association Statistics

	Mann-Whitney						
ROC Model	Area	Standard Error	95% Wald Confidence Limits		Somers' D (Gini)	Gamma	Tau-a
Hair	0.6110	0.0223	0.5672	0.6548	0.2220	0.3597	0.1055
Fingernail	0.6889	0.0218	0.6462	0.7315	0.3777	0.5399	0.1795

ROC Contrast Coefficients

ROC Model	Row1
Hair	-1
Fingernail	1

ROC Contrast Test Results

Contrast	DF	Chi-Square	Pr > ChiSq
Reference = Hair	1	10.4971	0.0012

With the help of Table 5, we can conclude that the difference between the AUCs for fingernail tests and hair tests is positive (0.0779) and that it has a standard error of 0.024. The Chi-square test results are repeated and we conclude that fingernail tests have the higher diagnostic accuracy than hair tests in depicting presence of alcohol abuse or dependence in this sample.

Table 5: Difference in AUCs

The LOGISTIC Procedure

ROC Contrast Estimation and Testing Results by Row

Contrast	Estimate	Standard Error	95% Wald Confidence Limits		Chi-Square	Pr > ChiSq
Fingernail - Hair	0.0779	0.0240	0.0308	0.1250	10.4971	0.0012

Word of caution: For specificity values above 0.92, the two ROC curves are virtually identical. Applying techniques and SAS® macros described in Shin (2009), it can be shown that we can be indifferent about the choice of biomarker if specificity values above 0.92 are desired since the sensitivity values are not statistically significantly different. Other decision criteria will have to take precedence in this case (see Zhou et al, 2012, for more details).

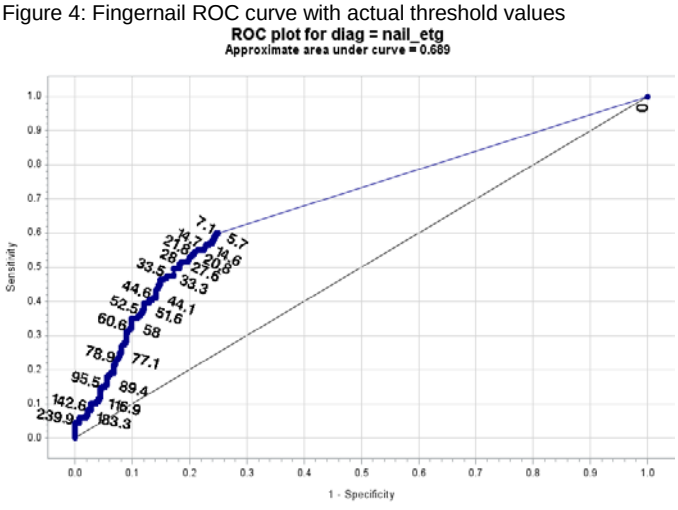
DETERMINATION OF OPTIMAL THRESHOLDS

For now, let us assume that we have decided to consider the fingernail test superior to the hair test. As a final step in the analysis, we would like to determine whether the original threshold value of 20 (pg/mg) of EtG was warranted. The following code imports the %rocplot SAS® macro and generates the same ROC curve as in Figure 2. The labels in Figure 4, however, correspond to actual values of the biomarker rather than predicted probabilities.

```
%inc "C:\Data\rocplot.sas";

ODS SELECT parameterestimates association;
PROC LOGISTIC DATA=etg_matched_sub;
       MODEL diag(EVENT='1') = nail_etg / OUTROC=roc1 ROCEPS=0;
       OUTPUT OUT=outp P=phat;
       ODS OUTPUT association=assoc;
run;
DATA _null_;
       SET assoc;
       IF label2='c' THEN CALL symput("area",cvalue2);
RUN;
TITLE "ROC plot for diag = nail_etg";
TITLE2 "Approximate area under curve = &area";
%rocplot(outroc = roc1,out = outp,p = phat,id = nail_etg,grid = yes)
```

Figure 4: Fingernail ROC curve with actual threshold values



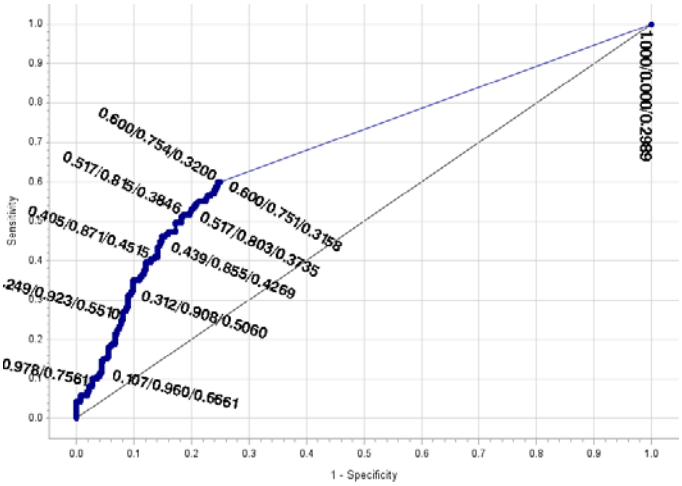
It appears that the threshold value just above 7 (pg/mg) has the largest distance from chance (vertical distance between 45-degree line and ROC curve) and the smallest distance from perfect prediction. The associated levels of sensitivity (0.60), specificity (0.754) and predicted probability (0.32) can be found in Figure 5. It was created with the

following code:

```
DATA outp;
SET outp;
pfmt=put(phat,6.4);
RUN;

%rocplot(outroc = roc1,out = outp,p = phat,id = _sens_ _spec_ pfmt,mindist=.05,grid =
yes)
```

Figure 5: Fingernail ROC curve with sensitivity, specificity and predicted probabilities



DISCUSSION

We used contingency tables and ROC analyses to determine whether fingernails were a more accurate biomarker than hair in detecting the presence of alcohol abuse or dependence in a sample of n=530 matched pairs from a large, random sample of college students. Based on unadjusted models, it appears that the fingernail test may have superior diagnostic accuracy than the hair test and that a lower cut-off value for that biomarker is warranted.

Study limitations include the risk for contamination of the physical specimens by chemical products (e.g. hair dye or perming), nail polish remover or hand lotions containing alcohol.

CONCLUSION

Several items included in the present analysis are going to be subject to change in the near future. The new Diagnostic and Statistical Manual of Mental Disorders (DSM-V) is scheduled for release in May 2013 and it will have one diagnosis of alcohol use disorder measured at four levels of severity (American Psychiatric Association, 2012). ROC models for more than two classes (see Krzanowski, 2009) should be employed in this context.

Zou et al (2012) suggest that rather than treating several biomarkers as competitors, diagnostic accuracy and AUC could be increased by pooling our hair and fingernail biomarkers. This is certainly an option we are going to explore in the near future.

Finally, one of the limitations of this study was that our models did not account for any other variables that could be predictive of alcohol abuse or dependence. Future research will make use of covariate-adjusted ROC models, compare biomarkers using the Lehmann family of ROC curves as a robustness check, and evaluate multivariate predictive models using ROC curves to address the issue of model over-fitting. Proper sample size and statistical power will be explored using techniques described in Mandrekar & Mandrekar (2005).

ACKNOWLEDGEMENTS

This work was sponsored by the National Institute on Alcohol Abuse and Alcoholism (NIAAA) grant R44AA016463. The funders did not play any role in study design, data collection and analysis, publication decisions, or the preparation of this manuscript.

REFERENCES

American Psychiatric Association (2000). Diagnostic and Statistical Manual of Mental Disorders: DSM-IV. American Psychiatric Association, Washington, DC.

American Psychiatric Association (2012). Proposed Revision: R 00 Alcohol Use Disorder. http://www.dsm5.org/ProposedRevisions/Pages/proposedrevision.aspx?rid=452# [accessed on August 15, 2012].

DeLong, Elizabeth R., D. M. DeLong, D.L. Clarke-Pearson (1988). Comparing the Areas Under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics*, 44, 837-845.

Gönen, Mithat (2006). Receiver Operating Characteristic (ROC) Curves. *Proceedings of the Thirty-first Annual SAS® Users Group International Conference*. http://www2.sas.com/proceedings/sugi31/210-31.pdf

Gönen, Mithat (2007). *Analyzing Receiver Operating Characteristic Curves with SAS®*. Cary, NC: SAS Institute Inc.

Krzanowski, Wojtek J., D. J. Hand (2009). *ROC Curves for Continuous Data*. Boca Raton: Chapman & Hall / CRC Press.

Jones, Joseph, M. Jones, C. Plate, D.Lewis, M. Fendrich, L. Berger, and D. Fuhrmann (2012). Liquid Chromatography-Tandem Mass Spectrometry Assay to Detect Ethyl Glucuronide in Human Fingernail: Comparison to Hair and Gender Differences. *American Journal of Analytical Chemistry*, 3, 83-91.

Lees, Rosemary, R. Kingston, T.M. Williams, G. Henderson, A. Lingford-Hughes, and M. Hickman (2012). Comparison of Ethyl Glucuronide in Hair with Self-Reported Alcohol Consumption. *Alcohol and Alcoholism*, 47(3), 267-272.

Mandrekar, Jay N. and S. J. Mandrekar (2005). Statistical Methods in Diagnostic Medicine using SAS® Software". *Proceedings of the Thirtieth Annual SAS® Users Group International Conference*. http://www2.sas.com/proceedings/sugi30/211-30.pdf

Obuchowsky, Nancy A., M. L. Lieber, and F. H. Wians, Jr. (2004). ROC Curves in Clinical Chemistry: Uses, Misuses and Possible Solutions. *Clinical Chemistry*, 50, 1118-1125.

Pepe, Margaret S. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*. New York, NY: Oxford University Press Inc.

Pragst, Fritz, M. Rothe, B. Moench, M. Hastedt, S. Herre and D. Simmert (2010). Combined Use of Fatty Acid Ethyl Esters and Ethyl Glucuronide in Hair for Diagnosis of Alcohol Abuse: Interpretation and Advantages. *Forensic Science International*, 196(1-3),101-110.

SAS Institute Inc. (2011). SAS/STAT® 9.3 User's Guide. Cary, NC: SAS Institute Inc.

Sheehan David, J. Janavas, K. Harnett-Sheehan, M. Sheehan, C. Gray, Y. Lecrubier, E. Weiller, T. Hergueta, C. Allgulander, N. Kadri, et al. (2009). Mini International Neuropsychiatric Interview. Tampa, USA.

Shin, Sanghyuk (2009). ROC analysis for the evaluation of continuous biomarkers: Existing tools and new features in SAS® 9.2. *Proceedings of PharmaSUG 2009 - Portland, OR*. http://pharmasug.org/download/papers/sp09.pdf

Society of Hair Testing (2011). Consensus of the Society of Hair Testing on Hair Testing for Chronic Excessive Alcohol Consumption 2011. http://soht.org

Zou, Kelly H., A. Liu, A. Bandos, L. Ohno-Machado, H. E. Rockette (2012). *Statistical Evaluation of Diagnostic Performance – Topics in ROC Analysis*. Boca Raton: CRC Press

Zhou, Xiao-Hua, N. A. Obuchowski, D. M. Obuchowski (2011). *Statistical Methods in Diagnostic Medicine* (2nd ed.). New York: John Wiley and Sons.

SAS® MACROS RESOURCES

SAS® Institute Inc. %ROCPLOT http://support.sas.com/kb/25/018.html

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Daniel Fuhrmann, Ph.D.
Center for Applied Behavioral Health Research, University of Wisconsin-Milwaukee
P.O. Box 786
Milwaukee, WI 53201
Work Phone: (414) 229-5176
Email: fuhrmann@uwm.edu
Web: http://www4.uwm.edu/cabhr/