medRxiv preprint doi: https://doi.org/10.1101/2020.05.12.20086868; this version posted May 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Use of Tablet-Based Neurocognitive Assessment Software for the Detection of Alcohol Impairment

Sanjeev Janarthanan[1*], Huy Phi[1*], Benjamin Flores[2], Yael Katz[2], David M. Eagleman[3], Bin Huang[2], Reza Hosseini Ghomi[2,4]

[1]College of Arts and Sciences, University of Washington, Seattle, WA, USA

[2]BrainCheck, Inc, Houston, TX, USA

[3]Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA, United States

[4]Department of Neurology, University of Washington, Seattle, WA, USA

[*]These authors contributed equally to this work.

Corresponding Author: B.H. (bin@braincheck.com) and R.H.G. (reza@braincheck.com).

Keywords

Alcohol Impairment, Neuropsychological Testing, Machine Learning, BrainCheck

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.12.20086868; this version posted May 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Abstract

Background: Acute ingestion of alcohol impairs cognitive function and poses significant threat to public health and safety with impaired operation of motor vehicles. However, there is a lack of access to tools to assess one's cognitive impairment due to alcohol. The purpose of this study was to explore the use of a neuropsychological assessment software, BrainCheck, to assess levels of alcohol impairment based on performance on the neuropsychological assessments.

Methods: We administered the BrainCheck battery to 91 volunteer participants. Participants were required to take a baseline battery prior to any alcohol ingestion, and another testing battery after a voluntary drinking period. Blood alcohol concentration (BAC) for the participant was obtained using a breathalyzer. We performed statistical analysis comparing alcohol vs. non-alcohol performance on the BrainCheck battery, and used significant metrics of these assessments to generate predictive models.

Results: Statistical analyses were performed comparing participants' performance on the BrainCheck battery before and after alcohol consumption. Comparison was also done comparing performance between an intoxicated group with a BAC > 0.05, and a sober group with a BAC ≤ 0.05. Two assessment metrics were found to be significant among comparison groups after P-value correction, and four test metrics were observed to moderately correlate ($|r| > 0.40$) with BAC levels. Three linear regression models (least-squares, ridge and LASSO) were built to predict participant BAC levels, with the best performing model being the least-squares model with a RMSE of 0.027. We also built a predictive logistic regression model to detect whether the participant is intoxicated or not, with 80.6% accuracy, 73.3% sensitivity, and 75.0% specificity.

Discussion: The BrainCheck battery has potential to predict alcohol impairment, including participant BAC levels and if the participant is intoxicated or not. BrainCheck provides another option to assess an individual's cognitive impairment due to alcohol, with the utility of being portable and available on one's smartphone.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.12.20086868; this version posted May 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Background

Alcohol is a widely known central nervous system (CNS) depressant, and acute consumption of alcohol results in impaired cognitive and psychomotor function, including reduced attention, alterations in memory, and reaction time [1]. The acute effects of alcohol pose significant threats to public health and safety. In 2018, 10,511 people died in alcohol-related motor vehicle accidents, accounting for 29% of all traffic-related deaths in the United States [2]. Although most legal limits for driving are set at 0.08 % blood alcohol concentration (BAC), research has shown that human subjects with BAC of 0.05 % display significantly diminished performance on psychomotor tasks of attention and reaction time compared to controls ($p < 0.05$) and even greater significance for BAC of 0.08% ($p < 0.01$) [3].

Conventional methods to detect alcohol impairment include the Standardized Field Sobriety Test which typically requires a police officer to administer the test at the roadside [4], however this method does not prevent drivers from operating their vehicles. Breathalyzers can also be used to calculate an individual's BAC, and although personal breathalyzers are becoming increasingly more accessible, they are still expensive, require calibration every 6-12 months, and lack consistency compared to police-grade breathalyzers [5]. Thus, a highly sensitive, rapid and self-administered cognitive screening test could aid in early detection of alcohol impairment, and prevent a intoxicated driver from operating a motor vehicle. Cognitive deficits associated with a BAC of 0.05% have been shown to be a reference for cognitive deficits associated with mild traumatic brain injury (mTBI) on neuropsychological assessment performance [6]. Prior work on computerized neuropsychological assessment batteries have also been able to assess impairment in cognitive functions with acute alcohol consumption [7], however this study was only used to assess performance on the neuropsychological assessment, and not in predicting the users level of alcohol impairment.

We explore the use of BrainCheck, a computerized neurocognitive assessment software that is available on iPad, iPhone or desktop browser. The BrainCheck testing battery administers neurocognitive tests, which work to maximize diagnostic accuracy, portability and ease of operator use. BrainCheck Sport has previously been validated for diagnostic accuracy for detection of mTBI [8], and BrainCheck Memory in identifying cognitive impairment and dementia [9]. Our aim is to test the utility of the BrainCheck battery on detecting acute alcohol impairment by comparing BrainCheck assessment composite scores with BAC obtained from participant breathalyzer scores.

Methods

Assessment selection

The BrainCheck battery contains five assessments, described in Table 1. Assessments are derived from traditional neuropsychological tests, including the Flanker test [10], Digit Symbol Substitution [11], Stroop [12], and the Trail Making A & B [13] tests. The coordination test is adapted from the Balance Error Scoring System [14].

medRxiv preprint doi: https://doi.org/10.1101/2020.05.12.20086868; this version posted May 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Assessment	Description	Measurement of Cognition
Flanker	We presented participants with a target arrow pointing to the left or right. The target was surrounded by congruent (> > > > >), or incongruent (<< > <<) arrows. Participants identified the direction of the target as quickly and accurately as possible.	Measures alertness, spatial awareness, and executive parts of attention
Digit Symbol Substitution	Participants must match an arbitrary correspondence of symbols to digits; when presented with a new symbol, they input as quickly as possible the corresponding digit.	Measures general cognitive performance
Stroop	Participants are instructed to find a word matching the given name of a color. There are two types of trials: CONGRUENT in which the word name and font color are the same (e.g., the word RED presented in red font), and INCONGRUENT in which the word name indicates a different color than the font (e.g., the word RED presented in green font).	Measures executive function and impulsivity
Trail Making A and B	Participants are instructed to connect a set of 25 dots in their correct order as rapidly as possible. Trail Making Test A uses only numbers (1 through 25), while Trail Making Test B employs alternating letters and numbers (1 – A – 2 – B – 3 – C - …)	Measures visual attention and cognitive flexibility
Coordination	A ball is displayed on the tablet, moving according to the tilt of the tablet. A participant holds the tablet out in front at arm's length, and tilts it appropriately to keep the ball in a central circle.	Measures balance and coordination ability

Table 1. Neuropsychological tests in the BrainCheck battery

Data collection

Inclusion criteria:

All participants required the use of both arms and legs, and perfect/corrected vision. Participants needed to be of legal drinking age (21), and have a breathalyzer BAC of 0.00 immediately prior to the first set of assessments.

Exclusion criteria:

We excluded participants with prior exposure to the battery, or those who admitted to alcohol or drug use within the previous 6 hours. We also excluded participants with impaired function of upper/lower extremities, memory disorders, imperfect/uncorrected vision or those who received less than 4 hours of sleep the previous night.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.12.20086868; this version posted May 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Obtaining participants:

To obtain participants, we approached likely candidates in public spaces and asked if they would volunteer for the study. Volunteers were collected through interoffice relationships, and random participants selected from a local pub in downtown Houston, Texas. Participants were required to take a baseline BrainCheck battery prior to any alcohol consumption, then took the battery a second time after an alcohol ingestion period of approximately 5-6 hours. Participants were not instructed or encouraged to consume alcohol. Only those who were consuming alcohol of their own volition regardless of their involvement in the study were asked to participate.

Statistical Analysis:

We used relative sample t-tests to compare assessment performance among participants for test 1 before alcohol consumption and test 2 after the alcohol consumption period. We used the Sidak correction method of t-tests for multiple comparisons to correct, and define a new significance value for $\boldsymbol{\alpha}$ [15]. Additionally, we calculated the difference of each assessment metric for each participant before and after alcohol consumption, and calculated the Pearson correlation coefficient (r) between them and the participant's BAC after consuming alcohol.

We also grouped all participants into two groups, those with a BAC above 0.05% comprised the intoxicated group, and those below or equal that threshold were in the sober groups. For each assessment, we again used relative sample t-tests and Sidak correction method to compare the assessment metrics of the intoxicated group with the sober group.

Next, we built linear regression models including least-squares regression, ridge regression, and least absolute shrinkage selection operator (LASSO) regression, to predict BAC based on the difference of assessment metrics before and after consuming alcohol. In order to minimize the impact of poor features and prevent overfitting, we applied L1 (Ridge) and L2 (LASSO) penalties to the model to restrict the feature weights. For the Ridge regression model, we tuned the L2 penalty from 0.0001 to 100,000, while we turned the L1 penalty from 0.0000001 to 10 in the LASSO regression model. We split 70% of the data into the training set, 10% into the validation set, and 20% into the test set. We used the validation data to evaluate the best L1 and L2 penalties for our LASSO and ridge models respectively.

We built Logistic Regression models with regularizations (L1, L2 and elastic net) to classify participants into the sober and intoxicated groups. We split 80% of the data into the training set and the other 20% into the testing set, based on the unique tester ID assigned to each participant. We evaluated the performance of the model using the receiver operator characteristic (ROC) curve and located the optimal threshold for the logistic model with maximum sensitivity/specificity.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.12.20086868; this version posted May 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Data analysis was performed using the Python programming language.

Results

Demographics:

We recruited 91 participants. This sample ended up being mostly individuals within the age range of 21-70 years old (47.3 % female). Participant demographics are displayed in Table 2. Mean BAC among all participants after the drinking period was 0.0997 (SD = 0.0373).

	Male/Female	**Age (Mean ± SD)**	**BAC after Drinking Period (Mean ± SD)**
Participants (n = 91)	48/43	34.37 ± 9.41	0.0997 ± 0.0373

Table 2. Participant Demographics

Assessment Metrics

All assessment metrics are displayed in Table 3. The Sidak method correction for multiple t-tests defined a new threshold for significant *p*-values, at $\boldsymbol{\alpha} = 0.005$. We found that metrics that are significantly different ($p < 0.005$) before and after alcohol consumption are the mean of trails A duration and median of trails A duration. Metrics that are significantly different for the intoxicated versus sober groups ($p < 0.005$) are the same. Boxplots that display the separation for metrics with a $p < 0.1$ are shown in Figure 1. We found four metrics (median of digit symbol duration, mean of digit symbols correct per second, mean of stroop reaction time and the median of incongruent stroop reaction time) with $|r| > 0.40$ (Figure 2), which would be considered as moderate correlations with BAC [16].

Assessment Metric	***P*-value for comparing Intoxicated vs. Sober groups**	***P*-value for comparing Test 1 vs. Test 2 groups**	**Pearson Correlation Coefficient (*r*)**	**Description**
Median of Digit Symbol Duration	0.16	0.19	**-0.45**	Total time taken to complete assessment. Time starts after digit display and ends when correct answer is completed

medRxiv preprint doi: https://doi.org/10.1101/2020.05.12.20086868; this version posted May 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Mean of Digit Symbols Correct per Second	0.20	0.26	**0.41**	Mean of correct responses in a second
Mean of Trails A Duration	**0.004**	**0.002**	-0.31	Average of response times between each click for Trail Making A
Mean of Trails B Duration	0.19	0.17	-0.36	Average of response times between each click for Trail Making B
Median of Trails A Duration	**0.0027**	**0.00005**	-0.10	Median of response times between each click for Trail Making A
Median of Trails B Duration	0.16	0.05	-0.25	Median of response times between each click for Trail Making B
Mean of Stroop Reaction Time	0.083	0.06	**-0.47**	Average reaction time for all responses
Median of Incongruent Stroop Reaction Time	0.096	0.04	**-0.44**	Median reaction time for all responses
Mean of Correct Flanker Reaction Time	0.072	0.08	-0.040	Median reaction time of all correct responses
Median of Correct Flanker Reaction Time	0.094	0.10	-0.17	Median reaction time of all responses when arrows point in the same direction
Mean of Balance Distance from Center	0.30	0.23	-0.22	Mean radial distance from center circle

medRxiv preprint doi: https://doi.org/10.1101/2020.05.12.20086868; this version posted May 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Table 3. Comparison of assessment metrics. *P*-values are from relative t-tests between sober vs intoxicated groups, intoxicated being BAC > 0.05, *p*-values comparing performance of participants before and after the drinking period, and *r* values are correlations between the change in the metric measurement and BAC levels. Bolded values are for *p*-values < 0.005 and |*r*| > 0.40.

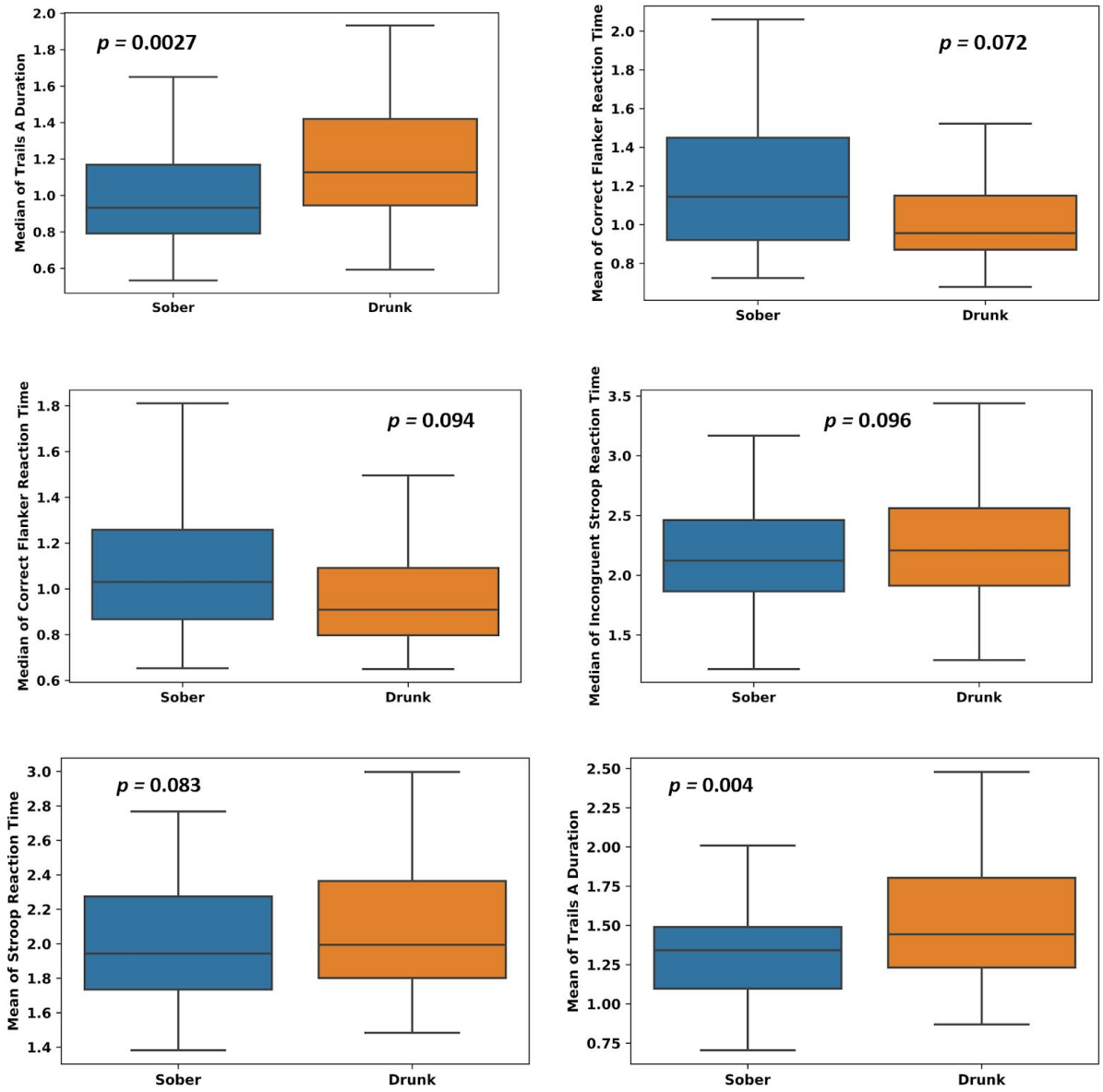


Figure 1. Comparison of assessment metrics between intoxicated and sober groups.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.12.20086868; this version posted May 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

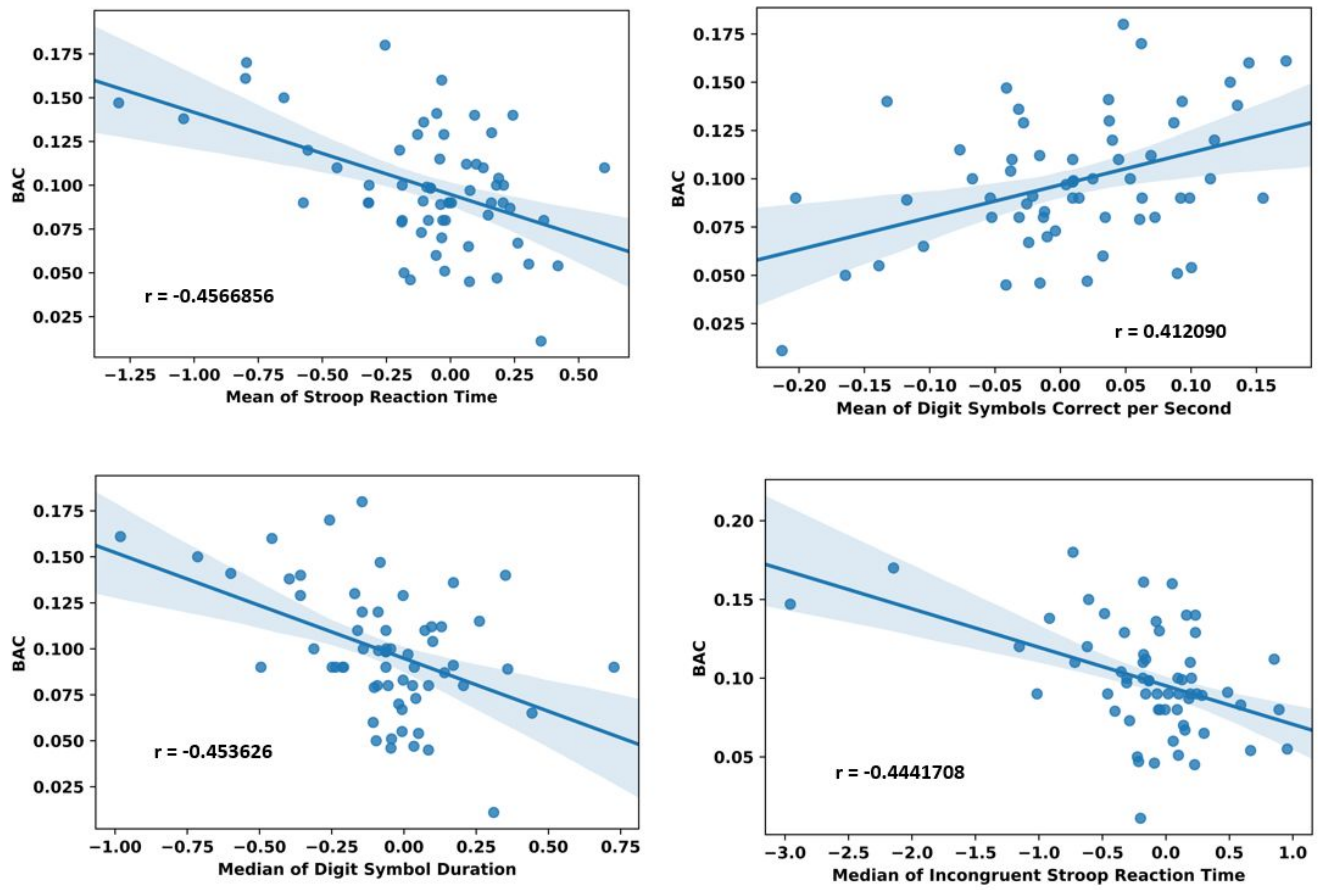


Figure 2. Correlations between the difference of assessment metrics before and after alcohol consumption and BAC.

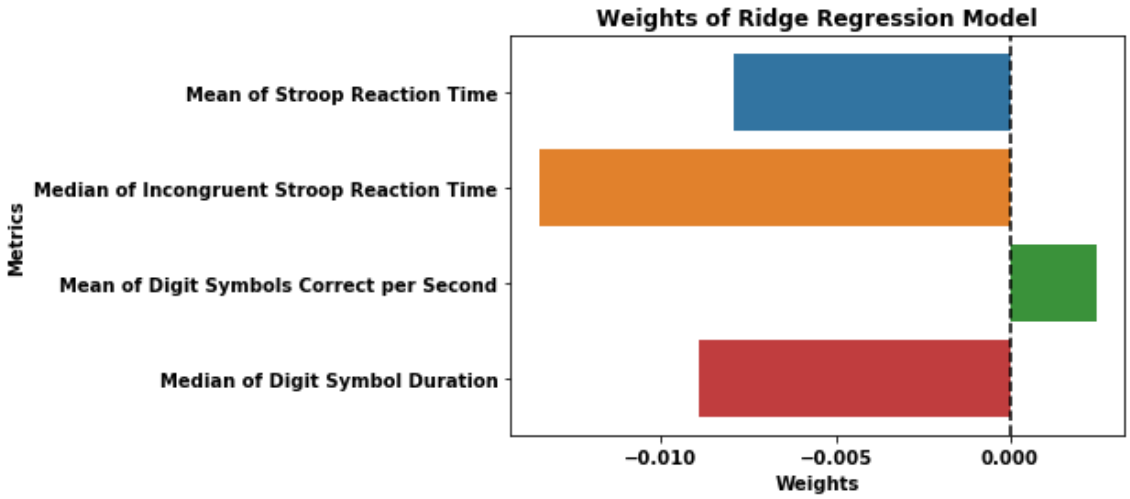
Linear Regression Models to Predict BAC

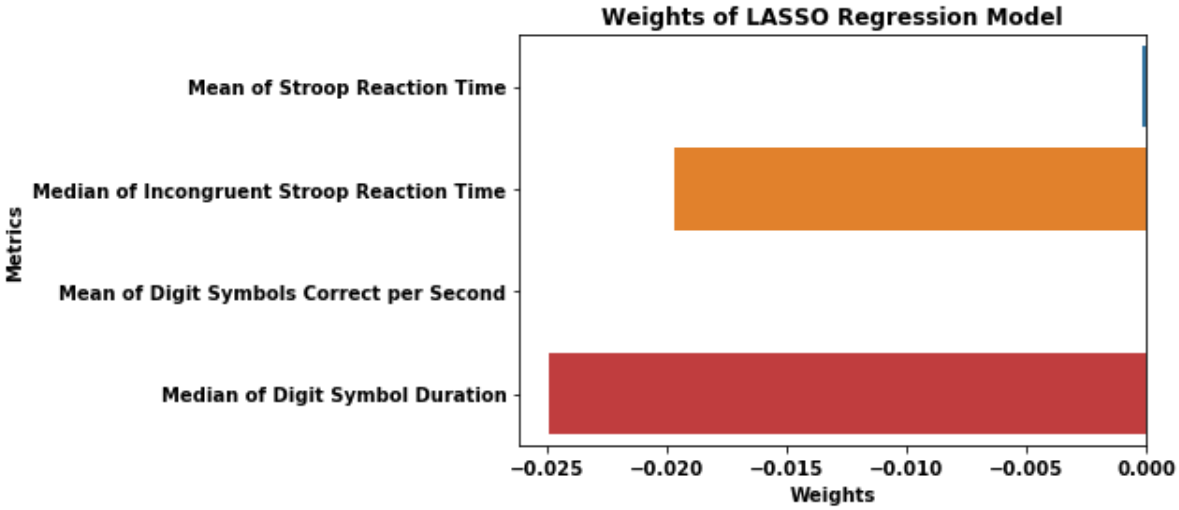
We used the metrics with $|r| > 0.40$ in Table 3 as input features for the linear regression models. We applied three linear regression models: least-squares regression, ridge regression, and LASSO regression to fit the training dataset, and evaluate their performance with Root Mean Square Error (RMSE). Their performance results are summarized in Table 4. The best model was the least-squares regression with a RMSE of 0.027. For reference, the mean BAC level measured at assessment, including before the drinking period, was 0.051 with a standard deviation of 0.057. We calculated a baseline RMSE, using the mean BAC as predicted values, which was 0.061. Our models demonstrate much better performance in predicting the actual BAC levels than the baseline method. For both the ridge regression and LASSO regression the features with the highest weights are the median of incongruent stroop reaction time and median of digit symbol duration. However, for the least-squares regression model, the features with highest weights are the mean of digits correct per second and median of digit symbol duration. The weights of the input parameters for these models are shown in Figure 3.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.12.20086868; this version posted May 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Table 4. Performance of Linear, Ridge and LASSO models in predicting BAC

Model	RMSE	L2 Penalty	L1 Penalty
Mean (baseline)	0.061301	--	--
Least-Square	0.026769	--	--
Ridge	0.030692	10	--
LASSO	0.029526	--	0.001





medRxiv preprint doi: https://doi.org/10.1101/2020.05.12.20086868; this version posted May 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

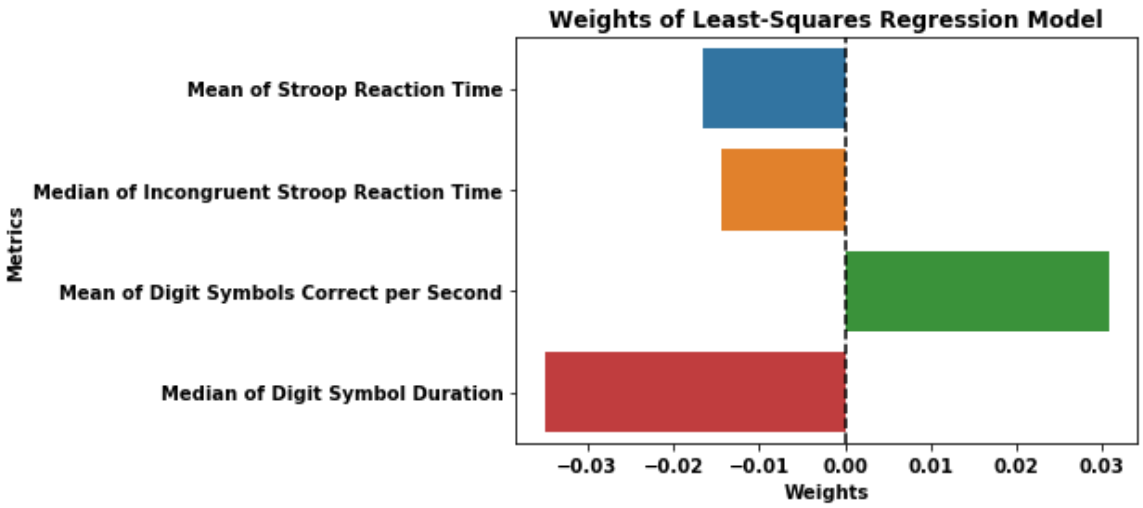


Figure 3. Weights of features for ridge, LASSO, and least-squares regression

Logistic Regression Model for Classification

We used metrics with the lowest *p*-value from each assessment type to build a logistic regression model with regularizations to predict whether participants were either sober or intoxicated. These metrics were the mean of balance distance from center, mean of stroop reaction time, median of trails A duration time, mean of correct flanker reaction time, and median of digit symbol duration. All participant scores were inputted into this model. We reported the recall, precision, and accuracy for different penalties in Table 5 below. We found that the logistic regression model performed the best with an L1 penalty. This model had an accuracy of 80.65 % and a precision of 0.91. Additionally, this model has nonzero weights for mean of trails a duration, median of trails a duration, and the mean of correct flanker time reaction, shown in Figure 4. We also calculated a ROC curve for this model on the test dataset. As shown in Figure 5, the optimal performance for this model was at a threshold of 0.46 with a sensitivity of 73.3 %, a specificity of 75.0 %, and an area under the curve (AUC) of 0.86. In other words, if the model predicted probability for a subject is greater than 0.46, the subject is categorized to the intoxicated group.

Table 5. Performance of the logistic regression model

Penalty	**Recall**	**Precision**	**Accuracy (%)**
L2	0.47	0.78	67.74
Elastic Net	0.80	0.50	51.61
L1	0.67	0.91	80.65

medRxiv preprint doi: https://doi.org/10.1101/2020.05.12.20086868; this version posted May 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

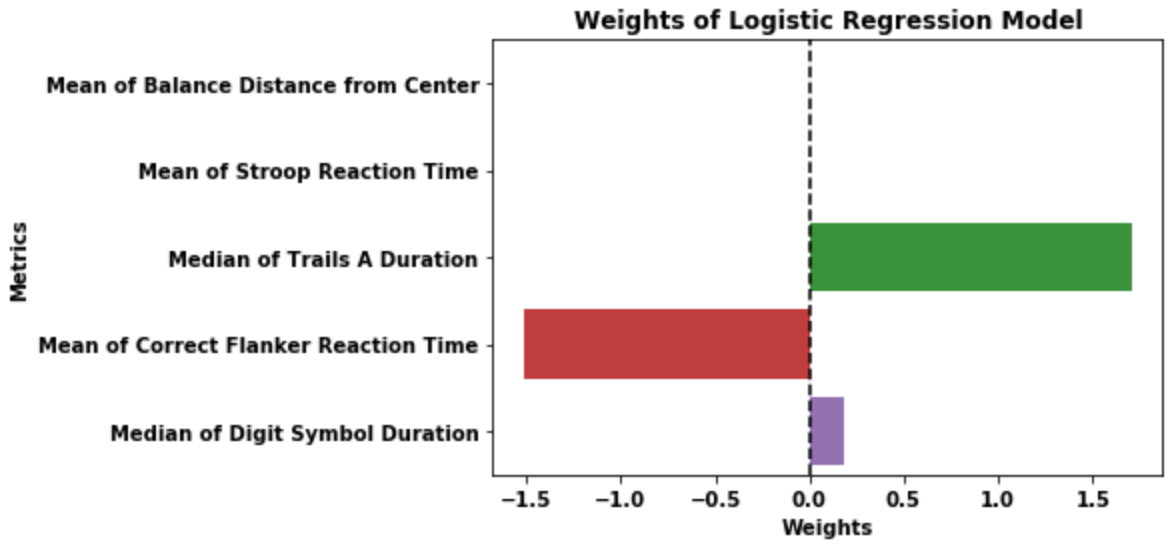


Figure 4. Weights of features for Logistic Regression Model with L1 Penalty

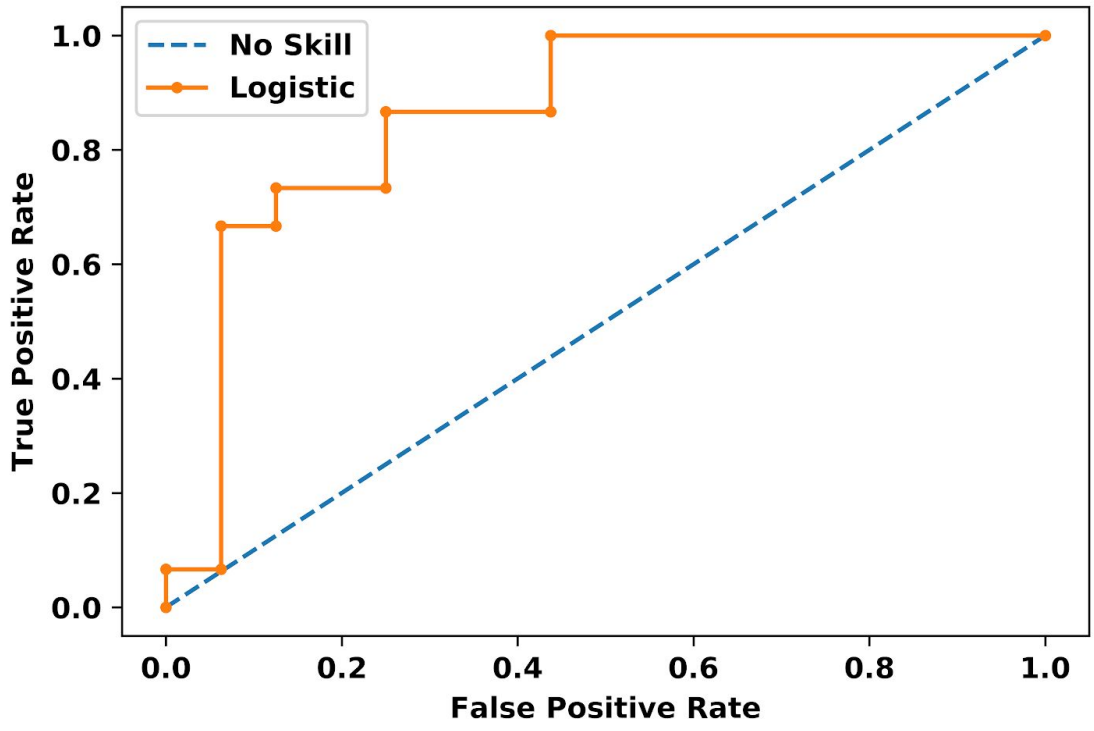


Figure 5. ROC Curve for logistic regression model with L1 penalty

medRxiv preprint doi: https://doi.org/10.1101/2020.05.12.20086868; this version posted May 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Discussion

The findings of this study demonstrated that the BrainCheck battery has acceptable levels of accuracy in predicting BAC and in classifying sober and intoxicated participants. The RMSE for our linear regression model was 0.027 which is significantly lower than the baseline RMSE of 0.061. In classifying intoxicated vs. sober participants, our best logistic regression model performed at an accuracy of 80.65%, with good sensitivity (73.3%) and specificity (75.0%). These results showed the potential of the BrainCheck battery to detect cognitive impairment under alcohol consumption, in a similar way to use the BrainCheck sport battery in detecting mTBI [8].

Our results also demonstrated the moderate correlations between the changes in cognitive performance with the consumption of alcohol. The correlation between performance changes on the digit symbol substitution assessment and BAC levels is consistent with prior work that has shown impaired performance on the digit symbol substitution test with acute consumption of alcohol [17]. Additionally, the correlations observed between reaction time in the stroop assessment and BAC levels are also consistent with impared performance in previous studies [18]. However, while we found significant differences in performance on Trails A, the previous work has observed only significant deficits in performance on Trails B, which is a more complex task than Trails A [19]. This different observation may be because we looked at an older population compared to the previous study (18-20 years old). The visuomotor performance defect by acute alcohol intoxication has been reported for older drinkers [20].

There are further limitations to discuss. First, the participants represent a convenience sample and the number of alcoholic drinks consumed were not standardised/consistent across all the participants, due to avoiding the encouragement of alcohol consumption purely for the study. Also, some assessments may not be sensitive enough for changes in mild alcohol intoxication. Average BAC in this study was relatively low, which may not be sufficient to cause cognitive impairment for most individuals. In addition, assessments were administered on the same day and used the individuals' baseline assessment as normal control, so we did not account for test-retest learning effects on participant assessment performance. Previous work on neuropsychological assessments demonstrate that participants generally perform better on the second assessment for both traditional and computerized tests [21] [22]. This learning effect could mitigate the declined performance caused by alcohol, which increases the challenge to distinguish the alcohol usage based on the battery performance.

Compared to other methods of assessing an individual's level of alcohol impairment, BrainCheck provides another option to assess level of alcohol impairment, including potential in predicting BAC and classifying intoxicated vs. sober. BrainCheck provides a unique option by providing a

medRxiv preprint doi: https://doi.org/10.1101/2020.05.12.20086868; this version posted May 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

shorter, gamified and portable test battery that can assess changes in cognitive function due to alcohol consumption, with the utility of being readily available on one's smartphone or tablet.

Conflicts of Interest

The following authors declare the following competing interests: BF, YK, BH and RHG reports personal fees from BrainCheck, outside the submitted work; BF, YK, DME and RHG reports receiving stock options from BrainCheck.

Acknowledgements

Study funding was provided by BrainCheck, Inc. SJ would like to thank the University of Washington Undergraduate Program in Neural Computation and Engineering for funding his work in preparing the manuscript. HP would like to thank the Mary Gates Endowment for providing a research scholarship for his work in preparing the manuscript.

Ethical Approval

This study protocol was reviewed and approved by Solutions IRB. All participants provided informed consent for being in the study.

Data Availability

Data may be made available by contacting the corresponding author and with a data use agreement.

Author Contributions

DME and YK contributed conception and design of the study; Data acquisition was performed by BF; SJ performed the statistical analysis, under the supervision of BH and RHG. HP wrote the first draft of the manuscript with assistance from SJ, which was revised and approved by BH and RHG.

References

[1] C. F. Valenzuela, “Alcohol and Neurotransmitter Interactions,” *RESEARCH WORLD*, vol. 21, no. 2, p. 5, 1997.

[2] NHTSA, “2018 Traffic Safety Facts: Alcohol Impaired Driving,” Dec. 2019. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812864 (accessed Apr. 28, 2020).

[3] S. A. Grant, K. Millar, and G. N. C. Kenny, “Blood alcohol concentration and psychomotor effects,” *British Journal of Anaesthesia*, vol. 85, no. 3, pp. 401–406, Sep. 2000, doi: 10.1093/bja/85.3.401.

[4] J. Stuster, *Validation of the Standardized Field Sobriety Test Battery at BACs Below 0.10 Percent: Final Report*. National Highway Traffic Safety Administration, 1998.

[5] B. C. Riordan, D. Scarf, S. Moradi, J. A. M. Flett, K. B. Carey, and T. S. Conner, “The accuracy and promise of personal breathalysers for research: Steps toward a cost-effective reliable measure of alcohol intoxication?,” *Digit Health*, vol. 3, Dec. 2017, doi:

medRxiv preprint doi: https://doi.org/10.1101/2020.05.12.20086868; this version posted May 18, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

10.1177/2055207617746752.

[6] A. Collie, P. Maruff, M. Makdissi, P. McCrory, K. Bennell, and D. Darby, "Quantifying the Cognitive Impairment Associated with Concussion: Using Blood Alcohol Concentration as a Reference Point," *Medicine & Science in Sports & Exercise*, vol. 34, no. 5, p. S94, May 2002.

[7] T. A. Schweizer, M. Vogel-Sprott, J. Danckert, E. A. Roy, A. Skakum, and C. E. Broderick, "Neuropsychological Profile of Acute Alcohol Intoxication during Ascending and Descending Blood Alcohol Concentrations," *Neuropsychopharmacology*, vol. 31, no. 6, pp. 1301–1309, Jun. 2006, doi: 10.1038/sj.npp.1300941.

[8] S. Yang *et al.*, "Diagnostic accuracy of tablet-based software for the detection of concussion," *PLoS One*, vol. 12, no. 7, Jul. 2017, doi: 10.1371/journal.pone.0179352.

[9] S. Groppell *et al.*, "A Rapid, Mobile Neurocognitive Screening Test to Aid in Identifying Cognitive Impairment and Dementia (BrainCheck): Cohort Study," *JMIR Aging*, vol. 2, no. 1, Mar. 2019, doi: 10.2196/12615.

[10] B. A. Eriksen and C. W. Eriksen, "Effects of noise letters upon the identification of a target letter in a nonsearch task," *Perception & Psychophysics*, vol. 16, no. 1, pp. 143–149, Jan. 1974, doi: 10.3758/BF03203267.

[11] J. Jaeger, "Digit Symbol Substitution Test," *J Clin Psychopharmacol*, vol. 38, no. 5, pp. 513–519, Oct. 2018, doi: 10.1097/JCP.0000000000000941.

[12] F. Scarpina and S. Tagini, "The Stroop Color and Word Test," *Front. Psychol.*, vol. 8, 2017, doi: 10.3389/fpsyg.2017.00557.

[13] C. R. Bowie and P. D. Harvey, "Administration and interpretation of the Trail Making Test," *Nature Protocols*, vol. 1, no. 5, pp. 2277–2281, Dec. 2006, doi: 10.1038/nprot.2006.390.

[14] D. R. Bell, K. M. Guskiewicz, M. A. Clark, and D. A. Padua, "Systematic Review of the Balance Error Scoring System," *Sports Health*, vol. 3, no. 3, pp. 287–295, May 2011, doi: 10.1177/1941738111403122.

[15] R. E. Blakesley *et al.*, "Comparisons of Methods for Multiple Hypothesis Testing in Neuropsychological Research," *Neuropsychology*, vol. 23, no. 2, pp. 255–264, Mar. 2009, doi: 10.1037/a0012850.

[16] P. Schober, C. Boer, and L. A. Schwarte, "Correlation Coefficients: Appropriate Use and Interpretation," *Anesthesia & Analgesia*, vol. 126, no. 5, pp. 1763–1768, May 2018, doi: 10.1213/ANE.0000000000002864.

[17] P. Thapar, J. P. Zacny, W. Thompson, and J. L. Apfelbaum, "Using Alcohol as a Standard to Assess the Degree of Impairment Induced by Sedative and Analgesic Drugs Used in Ambulatory Surgery," *Anesthesiology*, vol. 82, no. 1, pp. 53–59, Jan. 1995.

[18] I. A. David *et al.*, "Dynamics of a Stroop matching task: effect of alcohol and reversal with training," *Psychology & Neuroscience*, vol. 4, no. 2, pp. 279–283, 2011, doi: 10.3922/j.psns.2011.2.013.

[19] A. M. Day, M. A. Celio, S. A. Lisman, G. E. Johansen, and L. P. Spear, "Acute and Chronic Effects of Alcohol on Trail Making Test Performance Among Underage Drinkers in a Field Setting," *J Stud Alcohol Drugs*, vol. 74, no. 4, pp. 635–641, Jul. 2013.

[20] T. Brumback, D. Cao, and A. King, "Effects of Alcohol on Psychomotor Performance and Perceived Impairment in Heavy Binge Social Drinkers," *Drug Alcohol Depend*, vol. 91, no. 1, pp. 10–17, Nov. 2007, doi: 10.1016/j.drugalcdep.2007.04.013.

[21] W. R. Cole, J. P. Arrieux, K. Schwab, B. J. Ivins, F. M. Qashu, and S. C. Lewis, "Test–Retest Reliability of Four Computerized Neurocognitive Assessment Tools in an Active Duty Military Population," *Arch Clin Neuropsychol*, vol. 28, no. 7, pp. 732–742, Nov. 2013, doi: 10.1093/arclin/act040.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.12.20086868; this version posted May 18, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

[22] D. Hammers *et al.*, "Reliability of Repeated Cognitive Assessment of Dementia Using a Brief Computerized Battery," *Am J Alzheimers Dis Other Demen*, vol. 26, no. 4, pp. 326–333, Jun. 2011, doi: 10.1177/1533317511411907.