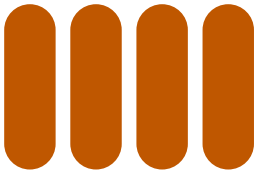
TURJ
Texas Undergraduate Research Journal
Volume 20
Fall 2021
texasurj.com

Depression and Alcohol Use

in Young Adults during the COVID-19 Pandemic

Page 93

Computer Science | Neurodevelopment | Psychology
Health Science | Medicinal Chemistry | Neuroscience



The University of Texas at Austin Undergraduate Research Journal is a student-run multidisciplinary journal of undergraduate research and an agency of the University Writing Center within the College of Liberal Arts

The University of Texas at Austin Undergraduate Research Journal

Inquiries should be addressed to:

editor.in.chief@texasurj.com
The University of Texas at Austin
Undergraduate Writing Center / URJ
2201 Speedway, Stop A6210
Austin, Texas 78712-1103
512-471-3166 | Fax 512-471-3408

Volume 20, Fall 2021
The UT Austin Undergraduate Research Journal is an agency of the
Undergraduate Writing Center. The University Writing Center is part of the Department of Rhetoric and Writing at The University of Texas at Austin.

Journal designs and layout by Nicholas Davila

On behalf of the 2020-21 editorial staff, I am honored to present the 20th volume of the Texas Undergraduate Research Journal (URJ). Throughout the past two decades, URJ has strived to encourage creativity, scholarship, and diversity in undergraduates at the University of Texas at Austin and showcase their talent in our annual publication. We also aim to increase the accessibility of science by holding networking events connecting students to opportunities, panel events to promote creative discussion, and research conferences for undergraduates to share their hard work. Over the past year, we expanded our mission to reach students in campuses across the state by accepting research manuscripts from institutions like Texas A&M and The University of Texas at Dallas. In addition, URJ co-leads the Capital of Texas Undergraduate Research Conference (CTURC) which connects students researchers in diverse disciplines from around Texas and hosts an academic space for undergraduates to practice presenting their research.

URJ's 2020-21 year was unorthodox given the shifts from COVID-19: all panels, events, conferences, and staff meetings were held virtually over Zoom. In light of the challenges presented by the pandemic, our latest edition would not have been possible without the dedication, hard work, and support from our authors, editors, and faculty reviewers. This issue features a selection of the best research - representing a rich diversity and depth of thought - that our university has to offer. We invite you to join us in exploring interdisciplinary work from how portrayal of healthcare communication on mass media impacts clinical trial patient recruitment to how virtual screening can be used to identify genes implicated in autism spectrum disorder.

It has been a privilege to work with the scientific talent at UT and across Texas. As readers, I hope you enjoy perusing this issue as much as we enjoyed producing it. I encourage you to explore our past volumes and student blog on our website, texasurj.com.

Sincerely,

Oishika Das
Executive Editor

FACULTY REVIEWERS

Eun Jeong Cho
Ingrid Edlund-Berry
Michael Domjan
Michael Mackert
Chris Beevers
Joseph Dunsmoor
Ioannis Giannikopoulos
Rabun Taylor
Elif Sarinay Cenik
Chris Whitman
Abimbola Adelakun
Antonio Linero
Kenneth Young
John R. Clarke
Gwen Stovall
Jennifer A. Miller
John Robert Bautista
Jarrod Lewis-Peacock
Juan M. Dominguez
Jiawei Han
David Taylor

ACKNOWLEDGEMENTS

The Texas Undergraduate Research Journal would like to extend special thanks to the Office of the Vice President for Research, the Office of Student Affairs, and the University Co-Op for their generous continued support of our mission throughout the Journal's years of operation.

Lastly, we would like to recognize the invaluable feedback of our faculty reviewers. And, most of all, we thank the authors themselves, whose inspiring research continues to change the world.

Journal Contents

ON THE COVER

Untitled, 8.5”X 11”
Mixed Media by Mei Robinson. The use of muted, darker colors refers to the the feelings of isolation and solitude permeating the world during the coronavirus pandemic. The masks scattered across the cover simultaneously promote a sense unity through orange symbolic to UT while aptly referencing the state of 2021. The black outlines evoke a sadness rooted in an emphasis on the negative aspects of life and pushing out of the positive aspects.

Classifier Performance Analysis through H-1B Visa Certification Status Classification

Yunhao Yang
Department of Computer Science, The University of Texas at Austin
Texas Undergraduate Research Journal

ABSTRACT

In 2019, US Citizenship and Immigration Services (USCIS) updated the H-1B application process. The new policy moves the H-1B lottery before the submission of the complete application. Therefore, the application certification status is more important since it becomes the final step before getting the visa. This paper explores the previous H-1B Application Dataset to find the patterns between applicants' information and their certification status. The paper demonstrates the correlations between several key features and the certification status. Moreover, it explores seven common classification models, analyzes the performance of each model according to the H-1B dataset, and integrates models in a specified way to achieve a 91% accuracy and 99% precision.[1]

1 BACKGROUND

The H-1B is a visa in the United States under the Immigration and Nationality Act, section 101(a)(15)(H) allows U.S. employers to temporarily employ foreign workers in specialty occupations. The application specifies the applicant's professional knowledge by requiring a bachelor's degree or equivalent work experience. The duration of the H-1B visa is between three to six years. After that, the visa holders need to reapply for the visa.

The process of getting an H-1B visa has three stages:

- The employer submits the United States Department of Labor a Labor Condition Application (LCA) for the employee, and provides relevant certifications, including a proof of wages (indicating that wages are at least equal to wages paid to other employees of the company in similar positions) and working conditions (Tai, 1991).
- With an approved Labor Condition Application, the employer files a Form I-129 (Petition for a Nonimmigrant Worker) requesting H-1B classification for the worker. This must be accompanied by the necessary supporting documents and fees (Seguritan, 2016).
- Once the Form I-129 is approved, the worker may begin working with the H-1B classification on or after the indicated start date of the job.

The research in this paper is focusing on the classification of the second stage, where the employee submits an application, supporting documents, and fees. The U.S. Citizenship and Immigration Services (USCIS) will review the documents and determine the case status for the application. The case status of any application can take one of four values:

[1] Code and datasets available at https://github.com/yunhaoyang234/H-1B_Data_Mining

- Certified: A certified Labor Condition Application (LCA), is a prerequisite to H-1B approval.
- Certified-Withdrawn: This means the LCA was approved but was later withdrawn by the employer, for some reason; It could be that the employee worked for some years before the contract was terminated.
- Denied: This means the LCA was denied; therefore, the necessary prerequisite for an H-1B approval is not in place.
- Withdrawn: This means the LCA was withdrawn before approval or denial. So, no decision was taken before the employer withdrew the application (George, 2018).

Historically, employers filed their full, and often voluminous, H-1B cap-subject petitions with USCIS, after which USCIS would select eligible petitions through a random selection process. This process resulted in unnecessary paperwork and incurred mailing costs for both petitioners and the agency (USCIS, 2019). Therefore, under the new policy, the applicants will go through the lottery pool first and submit their documents once they are selected. The new policy increases the importance of the certification status since it becomes the final step before getting the visa.

2 FACTS

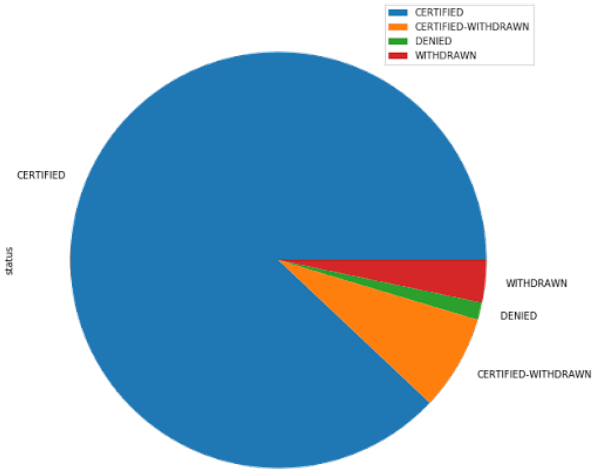


Figure 1: The figure shows the proportion of each visa application status. The blue, green, orange, and red parts represent Certified, Denied, Certified-Withdrew, and Withdrawn, respectively.

Based on the dataset of 2017-2018 H-1B Application Statistics, around 88 percent of applications were certified and moved to the next stage. The second-largest portion was Certified-Withdrawn, which comprises around 7.4 percent. Around 3.3 percent of applicants withdrew their application before it was reviewed. The last 1.3 percent of applications were denied.

2.1 Employer vs. Certified Rate

Table 1 shows the certified rate of large companies. In order to filter small companies out, I picked the companies with a number of applications greater than 5,000. The company with the highest certified rate is Infosys Limited, which has the largest application number as well. Technical companies contribute about half of the applications, such as Infosys, Microsoft, IBM, Google, etc. The majority of these companies have a certified rate above 90 percent. Moreover, Infosys, Tech Mahindra, and Tata Consultancy Services have certified rates above 99 percent.

Figure 2 is the number of applications of those companies sorted by the certified rate. From figure 2, we can see the overall trend is that the more applicants from the company, the higher the certified rate this company has. The number of applications from an employer has a positive correlation with the certified rate of this employer.

2.2 Worksite State vs. Certified Rate

Table 2 shows the top 10 states in the U.S. that have a large number of applicants and the highest certified rates. As shown in the table, although the application numbers are different, the certified rates of these states do not have much difference. Washington state has the highest certified rate - 90.72 percent. It is only 2 percent higher than the tenth place, Pennsylvania, which has a certified rate of 88.47 percent. Therefore, there is no clear correlation between the certified rate and location.

Company Name	Certified Rate
INFOSYS LIMITED	99.89
TECH MAHINDRA (AMERICAS),INC.	99.76
TATA CONSULTANCY SERVICES LIMITED	99.25
HCL AMERICA, INC.	98.87
ERNST & YOUNG U.S. LLP	98.84
ACCENTURE LLP	97.61
DELOITTE CONSULTING LLP	97.35
LARSEN & TOUBRO INFOTECH LIMITED	95.6
MICROSOFT CORPORATION	94.68
WIPRO LIMITED	92.62
CAPGEMINI AMERICA INC	92.35
COGNIZANT TECHNOLOGY SOLUTIONS US CORP	91.96
IBM INDIA PRIVATE LIMITED	80.04
GOOGLE LLC	78.23
IBM CORPORATION	73.42
GOOGLE INC.	62.81

Table 1

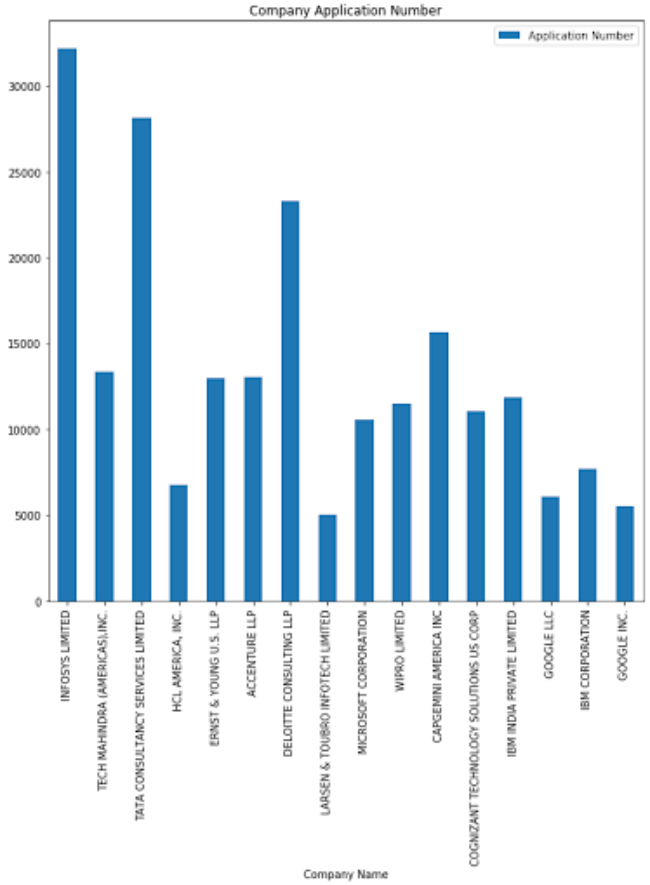


Figure 2: The x-axis lists the names of companies and the y-axis represents the number of applications in the corresponding company.

State	Application Number	Certified Rate
WA	52318	90.72
OH	31425	89.43
AZ	19482	89.17
TN	14126	89.17
WI	13190	88.95
NC	35625	88.86
MN	21299	88.81
IL	64041	88.78
GA	45832	88.75
PA	47428	88.47

Table 2

2.3 Job vs. Certified Rate

Computer-related jobs are generally easy to get certified. Table 3 shows the top 10 jobs with the highest certified rates. Seven of the top 10 job titles are computer-related, and their certified rates are all above 90%. Two of the top 10 jobs are related to business, and the last one is in the medical field. Nowadays, computer-related jobs are the most popular in and demanded by the country. Among these technical jobs, System Analyst and System Engineers have the highest certified rates, which are over 98%; system expertise is in high demand in the United States.

2.4 Salary vs. Certified Rate

Among all of the applicants for the H-1B visa, the majority have annual wages between $40,000 to $100,000. The mean of the applicants' salaries is $83,219, which is much higher than the real per-capita disposable income- $45,646 (FRB, 2019).

Figure 3 shows how the certified rate relates to the annual wage ranges. There is a strong positive correlation between wages and certified rates when the annual income is below $80,000. However, there is no clear correlation between wages and certified rates for wages higher than $80,000: the rates are fixed around 90%. The figure also indicates that applicants with annual wages lower than $40,000 have a relatively low chance to be certified

2.5 Applied Time vs. Certified Rate

Additionally, some applicants may care about how time relates to the visa-certified rate. There are two time-related features in the dataset, Application Submit Date and Employment End Date. It was found that the date that applicants submitted their application is not related to the certified rate. Figure 5 shows the month that applicants submitted their application versus the certified rate. The certified rates are all between 85% to 90%, and there is no increasing or decreasing pattern.

However, the year that applicants finish their employment is related to the certified rate. The data is organized by employment end year and demonstrates that the majority of the applicants will end their employment between 2017 ~ 2022 (the dataset is from 2017). There was a positive correlation between Employment End Year and the certified rate. The earlier applicants end their employment, the lower the chance their application will be certified (See figure 6). This observation was what we expected because companies, as well as the government, prefer employees who can work for a longer period of time. Longer employment associates with a greater contribution to the country.

Job	Application Number	Certified Rate
SYSTEMS ANALYST	31164	98.86
SYSTEMS ENGINEERS	2645	98.34
SOFTWARE TESTERS\t	4169	96.86
WEB DEVELOPERS	9492	96.5
INFORMATION SECURITY	5472	91.7
MANAGEMENT ANALYSTS	25955	90.72
NETWORK ARCHITECTS	3668	90.65
FINANCIAL SPECIALISTS	5626	90.44
COMPUTER OCCUPATIONS	114407	90.2
PHYSICAL THERAPISTS	5792	90.02

Table 3

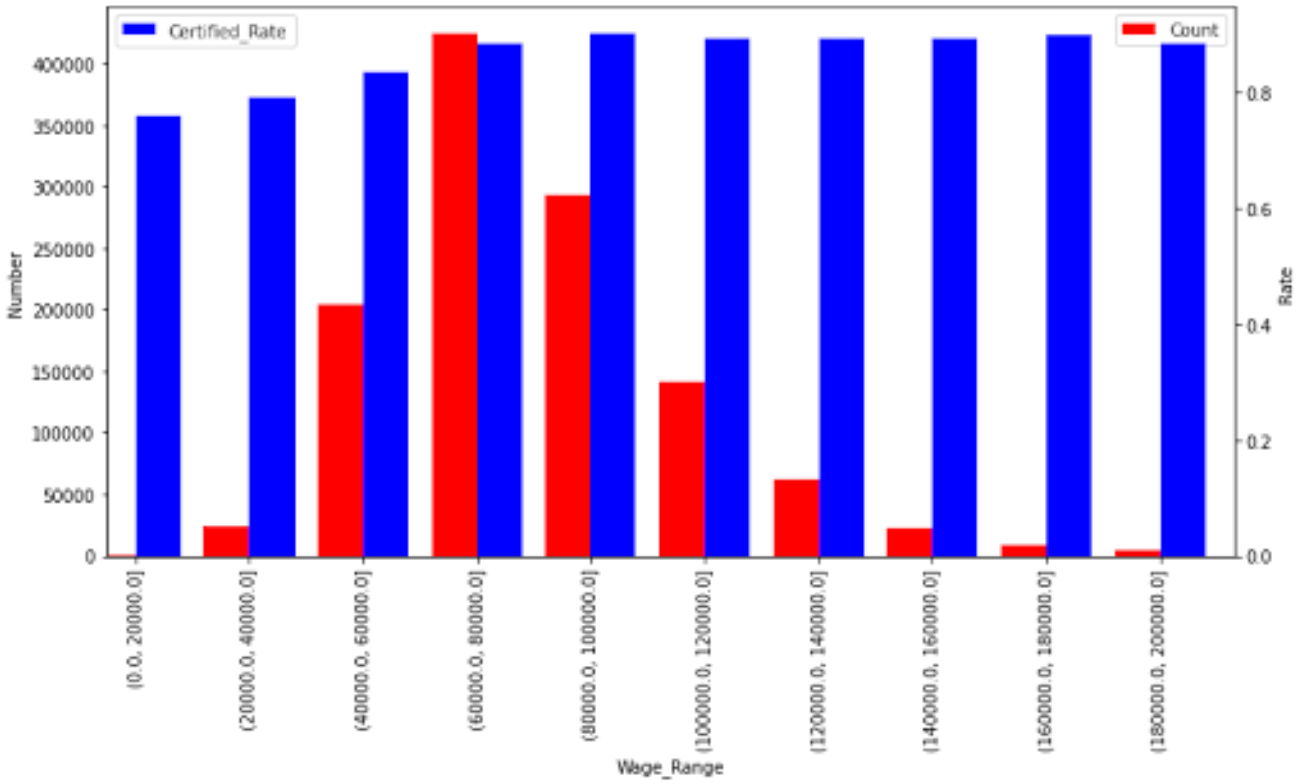


Figure 3: The figure shows the application number and certified rate on each range of salary. The x-axis represents the wage range of applicants. Each interval is 20,000 dollars. The left y-axis represents the number of applicants in a certain wage range. The right y-axis represents the certified rate.

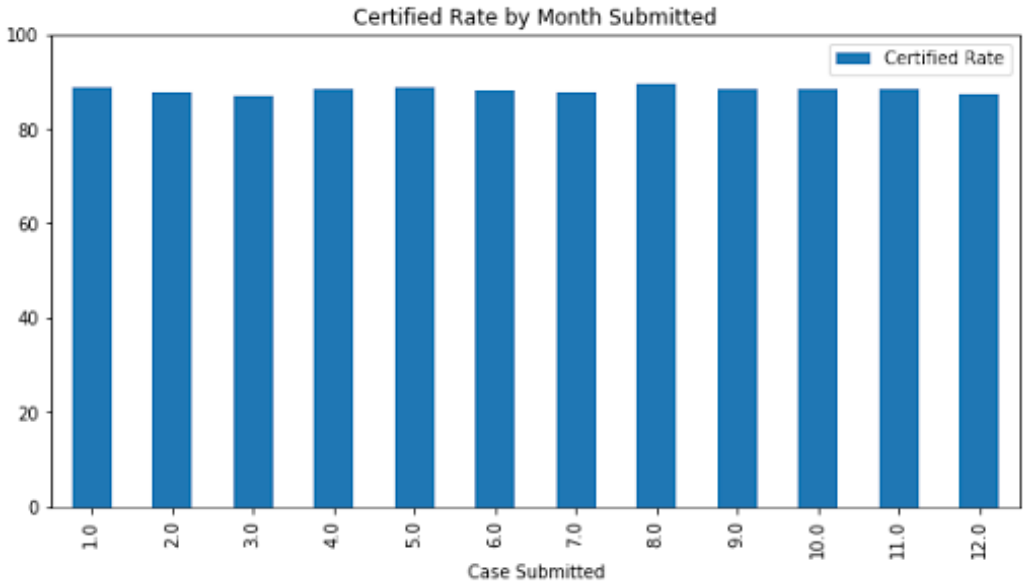


Figure 4: The x-axis represents the visa-certified rate, in percentage. The y-axis represents the number of applications submitted in a given month.

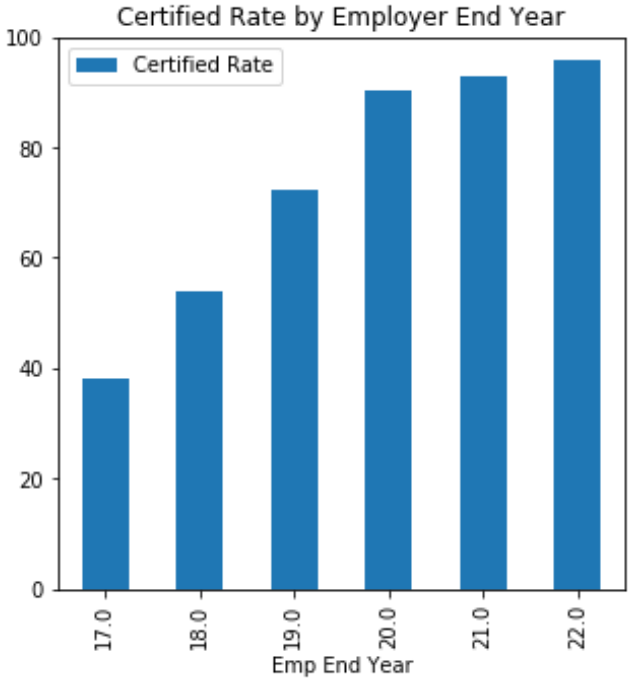


Figure 5: The x-axis represents the visa-certified rate in percentage. The y-axis represents the year that the applicant's contract with the current employer ends.

2.6 Employment End Year vs Certified Rate

However, the year that applicants finish their employment is related to the certified rate. The data is organized by employment end year and demonstrates that the majority of the applicants will end their employment between 2017 ~ 2022 (the dataset is from 2017). There was a positive correlation between Employment End Year and the certified rate. The earlier applicants end their employment, the lower the chance their application will be certified (See figure 6). This observation was what we expected because companies, as well as the government, prefer employees who can work for a longer period of time. Longer employment associates with a greater contribution to the country.

3 METHODOLOGY

3.1 Feature Engineering

3.1.1 Feature Selection

In the H-1B Application dataset, there are 50 features that specify information about the applicants. However, not all of the information is useful in classifying the certification status. For example, the CASE_NUMBER is a unique number that identifies the applicant, which is not related to the certification status. CASE_STATUS is the label that the models try to classify.

Section 2.2 and 2.5 show that the worksite state and application submission date do not affect the case status. In addition, there are multiple features that are highly correlated. The correlated features can be divided into the following three fields: Employer, Job, and Worksite. Therefore, only one feature in each field is selected because the other features in the field are associated with the selected feature.

Figures 7 and 8 are parts of the correlation matrix of features using the Pearson Correlation Coefficient.

This indicates employment_start_date and employment_end_date are highly correlated, as well as prevailing_wage and wage_rate_of_pay_from. As the correlation between two variables is too high, the two variables no longer convey additional information. Two or more highly correlated variables together will contribute to the determination of the classification status more than expected. Since two variables demonstrate the same set of attributes, these attributes will be considered twice during classification. Therefore, only one variable is kept in each group of highly correlated variables.

3.1.2 One-hot Encoding

Two features in the dataset are required to be one-hot encoded: Prevailing Wage Level and Prevailing Wage Source. Prevailing Wage Levels include from level 1 to level 4. The wage levels are defined as:

- Level I (entry) wage rates are assigned to job offers for beginning level employees who have only a basic understanding of the occupation.
- Level II (qualified) wage rates are assigned to job offers for qualified employees who have attained, either through education or experience, a good understanding of the occupation.
- Level III (experienced) wage rates are assigned to job offers for experienced employees who have a sound understanding of the occupation and have attained, either through education or experience, special skills or knowledge.
- Level IV (fully competent) wage rates are assigned to job offers for competent employees who have sufficient experience in the occupation to plan (ETA, 2009).

Prevailing Wage Sources include "Occupational Employment Statistics (OES)," "Collective Bargaining Agreements (CBA)," "Davis Bacon Act (DBA)," which will be used for any kind of construction-related work , "Service Contract Act (SCA)," and "Other."

3.1.3 Anomalies Removal

There are three features in the H-1B dataset that are numerical features with outliers:
Prevailing wage: annual wage for the job being requested for temporary labor conditions.
PW source year: year that the prevailing wage source was issued.
Wage rate pay to: maximum proposed wage rate.
Applying statistical anomalies detection using a normal distribution removed all of the records whose prevailing_wage, PW_source_year or Wage_rate_pay_to are beyond three standard deviations. For each feature, approximately five percent of the records are considered anomalies and removed from the dataset.

	EMPLOYMENT_START_DATE	EMPLOYMENT_END_DATE
EMPLOYMENT_START_DATE	1	0.849882
EMPLOYMENT_END_DATE	0.849882	1
TOTAL_WORKERS	0.014669	0.0351546
NEW_EMPLOYMENT	0.0251993	0.0407578
CONTINUED_EMPLOYMENT	0.0215066	0.0260076
CHANGE_PREVIOUS_EMPLOYMENT	0.0037496	0.013216

Figure 6: Two variables with high correlation will be marked red, and low correlation will be marked blue.

	PREVAILING_WAGE	PW_SOURCE_YEAR	WAGE_RATE_OF_PAY_FROM
EMPLOYMENT_START_DATE	0.00352235	0.146545	0.00321852
EMPLOYMENT_END_DATE	0.00466301	0.126674	0.00455481
TOTAL_WORKERS	0.000724649	0.00167033	-0.000422933
NEW_EMPLOYMENT	-0.000345467	-0.00342098	-0.00144235
CONTINUED_EMPLOYMENT	0.00115647	0.00366413	0.000868415
CHANGE_PREVIOUS_EMPLOYMENT	0.000630291	0.00307247	0.000244176
NEW_CONCURRENT_EMPLOYMENT	-0.00266486	0.00618635	-0.00317233
CHANGE_EMPLOYER	0.00203384	0.00569125	0.00200916
AMENDED_PETITION	0.000376054	0.00257725	-0.000164678
PREVAILING_WAGE	1	0.000800088	0.979902
PW_SOURCE_YEAR	0.000800088	1	0.000704215
WAGE_RATE_OF_PAY_FROM	0.979902	0.000704215	1
WAGE_RATE_OF_PAY_TO	0.00921472	-0.00171542	0.0104437

Figure 7: Continuation of Figure 6.

3.2 Metrics

Precision (also called positive predictive value) is the fraction of relevant instances among the retrieved instances. In this research, the most important metric is the precision of "denied" (label 0), which is the fraction of the number of correct predictions over the number of the applications that are predicted denied.

Recall (also known as sensitivity) is the fraction of the total amount of relevant instances that were actually retrieved. In this research, recall is the fraction of the number of correct predictions over the number of applications that are actually denied.

Accuracy is the overall percentage of the test records that were accurately predicted.

Among the three metrics, I am concentrating on precision. High precision indicates that if the model predicts rejection, the applicant is most likely to be rejected. Therefore, the applicants who get “denied” predictions may choose to give up this year and enhance their background for the next year. Thus, the predictions may help applicants save time or money that they spend on the applications. Moreover, a high precision will largely reduce the possibility of predicting rejection on someone who can actually be approved, which is the worst situation.

By contrast, if I am focusing on recall, the model will predict rejection on many applicants who may be able to get approval and cause them to give up their application.

3.3 Baseline Models

Training data: In order to avoid the effect of class imbalance, each model randomly chooses 50,000 records from the both the certified and the denied groups to form a balanced data with 100,000 records.

Testing data: In a random selection of 10,000 records, around 88% of the records are certified.

3.3.1 Decision Tree
Decision Trees are a non-parametric supervised learning method used for classification. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features.

The average cross-validation accuracy on balanced training data is 61.96%. By applying the classifier to the testing data, the classifier achieves a precision of 74% and an accuracy of 90%.

3.3.2 Random Forest
A random forest is a meta estimator that fits a number of decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting. The sub-sample size is always the same as the original input sample size but the samples are drawn with replacement.

The average cross-validation accuracy on balanced training data is 64.71%. By applying the classifier to the testing data, the classifier achieves a precision of 98% and an accuracy of 91%. By comparing to the decision tree, the precision significantly enhanced; while the accuracy also slightly increased.

3.3.3 AdaBoost
AdaBoost, short for Adaptive Boosting is a meta-estimator that begins by fitting a classifier on the original dataset and then fits additional copies of the classifier on the same dataset but where the weights of incorrectly classified instances are adjusted such that subsequent classifiers focus more on difficult cases.

The average cross-validation accuracy on balanced training data is 64.08%, which is slightly lower than the random forest classifier. By applying the classifier to the testing data, the classifier achieves a precision of 98% and an accuracy of 91%.

3.3.4 K Nearest Neighbors
Neighbors-based classification is a type of instance-based learning or non-generalizing learning: it does not attempt to construct a general internal model, but simply stores instances of the training data. Classification is computed from a simple majority vote of the nearest neighbors of each point: a query point is assigned the data class which has the most representatives within the nearest neighbors of the point.

The average cross-validation accuracy on balanced training data is 61.97%. By applying the classifier to the testing data, the classifier achieves a precision of 68% and an accuracy of 89%.

3.3.5 Support Vector Machine
Support-Vector Machines (SVM) build a model that assigns new examples to one category or the other, making them non-probabilistic binary linear classifiers. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on the side of the gap on which they fall (Cortes, 1995).

The average cross-validation accuracy on balanced training data is 64.04%. By applying the classifier to the testing data, the classifier achieves a precision of 92% and an accuracy of 90%.

3.3.6 Multi-Layer Perceptron (Neural Network)
Multi-layer Perceptron (MLP) is a supervised learning algorithm that learns a function f(*): Rm▸Rn by training on a dataset, where m is the number of dimensions for input and n is the number of dimensions for output.

Given a set of features $X=x_1,x_2,..., x_m$ and a target y, it can learn a nonlinear function approximator for either classification or regression. It is different from logistic regression, in that between the input and the output layer, there can be one or more non-linear layers, called hidden layers.

The average cross-validation accuracy on balanced training data is 63.47%. By applying the classifier to the testing data, the classifier achieves a precision of 92% and an accuracy of 89%.

3.4 Integration

There are two ensemble algorithms that can efficiently integrate the classification methods listed above.

3.4.1 Voting Classifier
Voting is one of the simplest ways of combining predictions from multiple machine learning algorithms. The voting classifier isn't an actual classifier but a wrapper for a set of different classifiers that are trained and evaluated in parallel in order to exploit the different peculiarities of each algorithm.

There are two ways of voting: 'hard' and 'soft' voting. The 'hard' method uses predicted class labels for majority rule voting. The 'soft' method predicts the class label based on the argmax of the sums of the predicted probabilities, which is recommended for an ensemble of well-calibrated classifiers. Based on the dataset, 'hard' method gives a better result.

The average cross-validation accuracy on balanced training data is 65.71%. By applying the classifier to the testing data, the classifier achieves a precision of 99% and an accuracy of 91%. This classifier achieves the highest precision, as well as accuracy.

3.4.2 Stacking Classifier
The simplest form of stacking can be described as an ensemble learning technique where the predictions of multiple classifiers (referred to as level-one classifiers) are used as new features to train a meta-classifier. The meta-classifier can be any classifier of your choice.

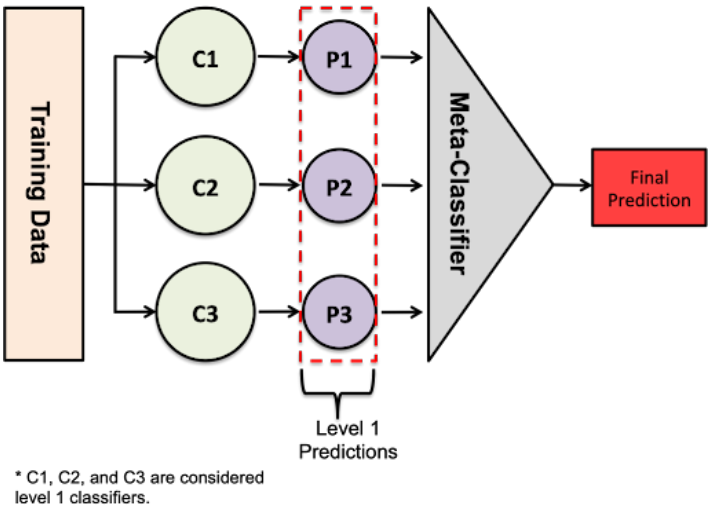


Figure 8: Architecture of stacking classifier.

The figure above shows how three different classifiers get trained. Their predictions get stacked and are used as features to train the meta-classifier which makes the final prediction. (Ceballos, 2019)

The average cross-validation accuracy on balanced training data is 62.10%. By applying the classifier to the testing data, the classifier achieves a precision of 49% and an accuracy of 88%. The stacking classifier does not achieve a satisfactory performance. This may be because of the inference between the level one classifiers.

4 RESULTS

4.1 Baseline Classifiers Evaluation

Considering the baseline classifiers' performance on testing data, the accuracies are between 88% to 91%. The majority of the classifiers get precision above 90%, except for the decision tree and nearest neighbor (around 70%). The recalls are below 30%. In terms of accuracy, two ensemble classification algorithms - Random Forest and AdaBoost - achieve the best accuracy, which is close to 91%. Decision Tree works best in terms of recall; Random Forest and AdaBoost achieve the highest precision which is 97.73%.
Among the accuracy, precision and recall, the most important metric is precision, which is the probability that the prediction is correct while the model predicts "denied (label 0)." The goal for the model is that if the model predicts "denied," this application has a high chance of being denied. So, the applicant can revise his/her application.

	Algorithm	Precision	Recall	Accuracy
0	decision tree	74.34	27.52	89.99
1	random forest	97.73	24.73	90.74
2	ada boost	97.73	24.73	90.74
3	nearest neighbor	68.32	18.02	88.97
4	svc	92.22	19.41	89.96
5	neural network	95.76	12.94	89.3
6	logistic regression	93.67	12.12	89.17

Table 4

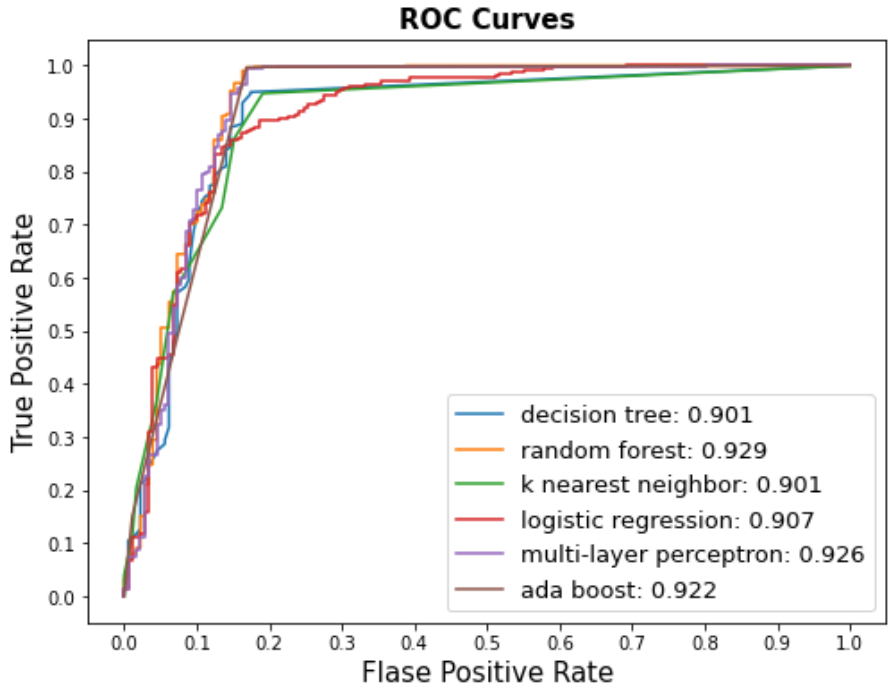


Figure 7: ROC Curves of some classifiers for the comparison of the performance. The scores in the bottom right labels are AUC scores.

The figure above shows the comparison of the classifiers' performance in terms of precision-recall tradeoff. The higher ROC curve indicates a better performance. Simultaneously, a higher ROC curve is corresponding to a higher AUC score. Since the ROC curves for the classifiers are close to each other, we can compare the AUC scores. From this figure, we can observe that the performance of all classifiers are close. However, random forest, ada-boost and multi-layer perceptrons achieve the highest three AUC scores, as well as the top 3 ROC curves.

Overall, the two ensembling methods Random Forest and AdaBoost have the best performance in terms of accuracy, precision, and recall.

	decision tree	random forest	ada boost	nearest neighbor	svc	neural network	logistic regression	label
decision tree	1	0.820695	0.820695	0.549303	0.664327	0.565078	0.539875	0.412865
random forest	0.820695	1	1	0.586125	0.781729	0.67547	0.644687	0.466451
ada boost	0.820695	1	1	0.586125	0.781729	0.67547	0.644687	0.466451
nearest neighbor	0.549303	0.586125	0.586125	1	0.718566	0.603376	0.553805	0.312621
svc	0.664327	0.781729	0.781729	0.718566	1	0.747898	0.689045	0.396892
neural network	0.565078	0.67547	0.67547	0.603376	0.747898	1	0.871194	0.330528
logistic regression	0.539875	0.644687	0.644687	0.553805	0.689045	0.871194	1	0.31525
label	0.412865	0.466451	0.466451	0.312621	0.396892	0.330528	0.31525	1

Figure 10

4.2 Similarity Studies

Figure 10 is the similarity matrix indicating the similarity between the outcomes of classifiers and the correct label. Coincidentally, Random Forest and AdaBoost give the exact same outputs. In addition, Decision Tree has a high correlation with Random Forest and AdaBoost, which is understandable since both Random Forest and AdaBoost use multiple decision trees. Neural Networks (Multi-Layer Perceptron) and Logistic Regression are highly correlated as well.

In general, all the similarities between classifiers are higher than the similarities between labels and classifiers, which indicates all the classification algorithms tend to generate similar outcomes. Therefore, if one classifier predicts a result, the other classifiers are likely to predict the same result even though it is not the correct label.

4.3 Integration results

There are two simple classification algorithms that are used to integrate the results of other classifiers, Voting Classifier and Stacking Classifier.

The performance of these classification methods is shown in Table 5. The Stacking Classifier does not achieve significantly higher accuracy and precision, but it reaches around 29% recall, which is significantly higher than other classifiers. By contrast, the Voting Classifier does significantly improve the performance.

The precision of the Voting Classifier is 99.23%, which is 1.5% higher than the Random Forest and AdaBoost (baseline algorithms that get the highest precision). The Voting Classifier gets 21.49% on recall, which is the median of the nine classifiers. The Voting Classifier achieves the third highest accuracy, only 0.15% below the Random Forest and AdaBoost.

Algorithm	Precision	Recall	Accuracy
decision tree	74.34	27.52	89.99
random forest	97.73	24.73	90.74
ada boost	97.73	24.73	90.74
nearest neighbor	68.32	18.02	88.97
svc	92.22	19.41	89.96
neural network	95.76	12.94	89.3
logistic regression	93.67	12.12	89.17
voting	99.23	21.49	90.59
stacking	48.66	28.93	87.85

Table 5

5 CONCLUSIONS

5.1 Final Model

The research develops a way of classifying H-1B visa certification status, which achieves 91% accuracy, 99% precision, as well as almost 21% recall.

This model is designed to maximize precision. A 99% precision means if the model predicts this application will be denied, then this application has a 99% chance of being denied. So, the applicant knows that the application needs to be revised and improved. Accuracy and recall are not the top priority for this model.

The final model is constructed by seven distinct classifiers listed in section 3.3. Each classifier is tuned independently. Then, Voting Classifier is used to integrate the seven trained classifiers and tuned to fit the training data.

5.2 Classification Algorithm Analysis

The outcomes of the classification methods in section 4.1 and 4.2 demonstrate two facts about the classification algorithms:

- Ensemble classifiers have a significant improvement in performance compared to other classifiers. In thise research, Random Forest and AdaBoost achieve the best performance among seven models.

- All the models tend to make similar predictions. Although the classification algorithms are different, the similarity between the predictions is typically higher than the similarity between predictions and actual labels. Therefore, even though combining multiple classifiers may improve the performance (accuracy, precision, and recall), the improvement will not be significant.

ACKNOWLEDGMENTS

Thanks to Prof. Angela Beasley, Assistant Professor of Instruction in the Department of Computer Science - University of Texas at Austin, for the supervision and guidance throughout this project.

REFERENCES

Tai, Wendy S. (September 29, 1991). "'Prevailing wage' has high-tech companies on edge." *Star Tribune* (Minneapolis, Minnesota). p. 19A.

Seguritan, Reuben S, Esq. (March 25, 2016). "USCIS Issues H-1B Filing Instructions." *The Filipino Express* (Jersey City, New Jersey). p. 10, 12.

George, Olawunmi. (Apr 29, 2018). The H-1B: An analysis of American companies' requests for external labor. https://codeburst.io/the-h-1b-an-analysis-of-american-companies-requests-for-external-labour-2f9fc718613b

"Real Disposable Personal Income: Per Capita [A229RX0]." *Federal Reserve Bank of St. Louis. U.S. Bureau of Economic Analysis.*

USCIS Announces Implementation of H-1B Electronic Registration Process for the Fiscal Year 2021 Cap Season (Dec. 6, 2019). *US Citizenship and Immigration Service.*

Employment and Training Administration. (Nov 2009). *Prevailing Wage Determination Policy Guidance, Nonagricultural Immigration Programs.*

Cortes, Corinna; Vapnik, Vladimir N. (1995). "Support-vector networks" (PDF). *Machine Learning*. 20 (3): 273–297. CiteSeerX 10.1.1.15.9362. DOI:10.1007/BF00994018.

Tolles, Juliana; Meurer, William J (2016). "Logistic Regression Relating Patient Characteristics to Outcomes." JAMA. 316 (5): 533–4. DOI:10.1001/jama.2016.7653. ISSN 0098-7484. OCLC 6823603312. PMID 27483067.

Ceballos, Frank. (July 20, 2019). "Stacking Classifiers for Higher Predictive Performance." *towarddatascience.com* https://towardsdatascience.com/stacking-classifiers-for-higher-predictive-performance-566f963e4840

Pedregosa, F. Varoquaux, G. Gramfort, A. Michel, V. Thirion, B. Grisel, O. Blondel, M. Prettenhofer, P. Weiss, R. Dubourg, V. Vanderplas, J. Passos, A. Cournapeau, D. Brucher, M. Perrot, M. Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python, *Pedregosa et al.*, JMLR 12, pp. 2825-2830.

Status of Fictional Televised Clinical Trial Communication by Healthcare Practitioners

Jacey Pridgen
College of Natural Science Polymathic Honors Programs, The University of Texas at Austin
Texas Undergraduate Research Journal

ABSTRACT

Many clinical trials fail due to low patient participation and recruitment, often as a result of poor health care practitioner (HCP) to patient communication. Furthermore, mass media can have widespread effects on viewers' behaviors. Previous work has examined the content of fictional clinical trial representations in television from a sociological perspective. However, the content within televised portrayals of HCP and patient interactions from a communication perspective has not been studied. The purpose of this study is to describe the status of televised fictional clinical trial communication by HCPs to explore potential barriers related to participant recruitment. By analyzing fictional HCP and patient communication within a clinical trial scenario, I aimed to explore possible influences that deter clinical trial participation. I completed a content analysis of recent television medical dramas, measuring both quantitative and qualitative factors of HCP communication. Overall, I found that the fictional communication represented within analyzed television dramas was patient-centered, empathetic, and categorically positive. These findings suggest that fictional televised medical dramas most likely do not act as a further barrier to clinical trial recruitment. This study illustrated the opportunity for health communications working with the media community to better tailor fictional communication to improve messaging about clinical trials and increase recruitment.

INTRODUCTION

Clinical trials and medical research hold a position of vast importance for stakeholders in the pharmaceutical industry, healthcare practitioners (HCPs), and all other people who utilize modern medications and procedures. Novel pharmaceuticals and treatments must first be tested in a scientifically rigorous manner before approval for safe usage (Novitzke, 2008). The last stages of testing require human participants who are recruited to the clinical trial (NIA, 2017). However, many clinical trials are unsuccessful because they do not have enough recruits (Fogel, 2018). For example, two recent reviews of clinical trials based in the United States found that low recruitment was the leading cause of clinical trial failure (Stensland et al., 2014; Nguyen et al., 2018). In the United States, where the median research and development costs to bring a new pharmaceutical to market is $985 million, it is important not to incur unnecessary costs through clinical trial failures (Wouters et al., 2020).

The failure of clinical trials to attract recruits is due to a variety of reasons stemming from structural, demographic and socioeconomic, clinical, and attitudinal barriers (Unger et al., 2016). Demographic and socioeconomic, structural, and clinical barriers have been studied to a greater extent than attitudinal barriers. Attitudinal barriers deserve more considerable attention from researchers as they are purely patient held, rather than a condition of the patient's environment. This is significant as patient-held beliefs can potentially be reshaped by healthcare practi-

tioners while the patient's societal environment is much harder to control.

It is known that media can have widespread influences on viewers' behavior. Several studies have implicated media as a significant shaper of the media consumer's viewpoints, opinions, and behaviors (Rideout, 2008; Dudo et al., 2011; Hether et al., 2008; Quick, 2009). Because the media can influence behavior, it calls into question if the media may influence whether or not people join clinical trials. To study this phenomenon, Fisher and Cottingham (2017) conducted a content analysis of fictional representations of clinical trials within television episodes and films. They found that fictional clinical trials are portrayed negatively overall, though affirmative consent was depicted in many instances (Fisher and Cottingham, 2017). Their perspectives as sociologists may have colored these findings as they were not explicitly analyzing communication events. Exploring the fictional portrayals of clinical trials through a communication lens may lead to new and interesting findings.

The relationship between an HCP and a patient is primarily forged through successful communication. (Ha et al., 2010). Patient-centered communication has become the hallmark of ideal healthcare practitioner communication globally (Hashim, 2017). The Institute of Medicine lists patient-centered communication as one of the six indications of model patient care (2001). The Institute describes patient-centered communication as "providing care that is respectful of and responsive to individual patient preferences, needs, and values, and ensuring that patient values guide all clinical decisions" (pg. 3). Though strategies for patient communication must be variable and flexible, if HCPs utilize patient-centered communication and shared decision making as a guideline for care, both HCPs and patients will achieve their goals (Levit et al., 2013).

Communication is the primary mediator of the HCP and patient relationship. This, coupled with the fact that the impact of mass media on people joining clinical trials is unknown, raises the question of how fictional representations of HCP communication regarding clinical trials in mass media could influence potential clinical trial recruit behavior. It is first necessary to survey the current representations of HCP communication efforts in fictional media before directly measuring patient perceptions of clinical trials after viewing fictional HCP conversations in fictional media.

Though a general content analysis of clinical trial representations in fictional media was completed by Fisher and Cottingham in 2017, it is necessary to further their analysis. By specifically analyzing fictional HCP and patient communication within a televised clinical trial scenario, I will more accurately detail whether positive, neutral, or negative communication events are being portrayed. To determine whether the interactions are similar to typical HCP and patient communication events, I will also describe the quantitative traits of these interactions, such as speaking times and numbers of questions asked by the HCP and patient. This study is an important first step to investigate possible fictional media influences that could impact clinical trial participation.

BACKGROUND

Clinical Trials

As defined by the National Institutes of Aging (NIA) within the National Institutes of Health, "clinical trials are research studies performed in people that are aimed at evaluating a medical, surgical, or behavioral intervention" (2017, para. 3). Clinical trials have four phases. Only Phase I, II, and III trials are necessary for the FDA to approve a novel treatment. Phase I trials test new treatments in a small population of individuals to explore the side effects, safety, and to establish the proper dosages of the treatment. Phase II trials are done in larger populations to determine if the treatment is effective in the intended population, as well as to further explore the safety and side effects. Phase III trials are the last step before FDA approval. In a Phase III trial, the novel drug is examined within different groups of individuals and in combination with different drugs. This trial is done with a large number of people and advances the researchers' knowledge of safety and effectiveness compared to other treatments. If the trial has positive results, then the FDA could approve the treatments.

Phase IV trials can be optional, generally examining the long-term impacts of the drug or treatment (NIA, 2017).

Benefits and harms

While clinical trials are useful to advance science overall, they also directly benefit specific populations. A proportion of clinical trial participants are healthy volunteers—people without a specific medical condition that test a pharmaceutical to determine the side effects (Stunkel & Grady, 2011). These people are financially compensated and have reported other motivations for participating, such as curiosity, interest in science, and wanting to help further science and health (Stunkel & Grady, 2017). The other cohort of clinical trial participants are patients who have the disease or condition the novel treatment is thought to treat (NIA, 2017). These patients report a variety of benefits. For example, a 2008 study of patients enrolled in cardiac disease clinical trials found that the hospitals enrolling patients in clinical trials implemented higher quality of care and had reduced mortality rates compared to non-enrolling hospitals (Majumdar et al., 2008). A 2001 review of published clinical trial primary data further revealed that clinical trials had a weakly positive impact on patient outcomes (Braunholz et al., 2011) The positive outcome was also furthered when the trial treatment was controlled with a known effective treatment (Braunholz et al., 2001). In the third example of patients benefiting after enrolling in clinical trials, patients with glioblastoma were found to have significantly increased survival advantage over patients not enrolled in studies (Sharar et al., 2012). Overall, patients in clinical trials can benefit in various ways.

There are many direct benefits for patients in clinical trials, but there can also be substantial harm. Though not the primary reason for clinical trial failure, trials of any phase can fail due to adverse side effects or safety issues. Hwang et al. completed a 2016 review of Phase III that analyzed trials between 1998 and 2008 and found that 17% failed due to safety issues. These safety issues ranged from "serious adverse effects such as cancer, stroke, and sepsis" to "increased risk of death" (Hwang et al., 2016, pg 1829). For example, Kaur et al. (2016) detailed the failed BIA 10-2474 Phase I trial. In this trial, four volunteers had irreversible brain damage, and one patient was declared brain dead due to a novel fatty acid amide hydrolase inhibitor. Attarwala (2010) details the failures within the TGN1412 Phase I clinical trial. In this trial, six healthy volunteers were given a novel CD28 superagonist antibody, and soon after were transferred to an intensive care unit due to detrimental complications such as multiorgan failure (Attarwala, 2010). These scenarios show that while clinical trials overall have bettered patient outcomes and furthered science, individual recruits are the ones being impacted more directly.

Problems

Clinical trials are significant as all novel pharmaceuticals and treatments must first be tested in a scientifically rigorous manner before being approved for general usage (Novitzke, 2008). A significant problem that arises is that many clinical trials are not completed due to difficulty with recruitment and accrual (Fogel, 2018). For example, recent figures show that only 3% to 5% of cancer patients elect to join cancer clinical trials, leading to low numbers of recruits for the most common types of clinical trial (Unger et al., 2016; Collins, 2011). The low number of patients joining clinical trials can lead to trial failure. Bounansegna et al., (2014) interviewed pharmaceutical industry leaders to explore their expert opinions regarding non-drug related failure of clinical trials. The industry leaders listed recruitment issues as one of the primary non-drug related failures (Buonansegna et al., 2014). Furthermore, a review of clinical trials registered on the U.S. ClinicalTrials.gov webpage done by Stensland et al. (2014) found the largest proportion of clinical trials, 39%, failed due to low accrual numbers. In a separate review of radiation therapy trials on U.S. ClinicalTrials.gov, Nguyen et al. (2008) also found that 57.5% of trials failed due to a lack of recruits.

The second problem is that many clinical trials lack generalizability due to the underrepresentation of specific populations (Stuart et al., 2015). There are differences in the health status of people who join clinical trials when compared to people who do not join clinical trials. For example, Elting et al. (2006) found that people with cancer enrolled in clinical

trials were younger and had fewer comorbid conditions than non-enrolled people with cancer. In contrast, Elting et al. (2006) also found that enrolled people with cancer were more likely to have more advanced cancer compared to non-enrolled people with cancer. These differences in trial populations lead to trials which cannot always be generalized across the entire population of people with a specific disease or disorder. There are also racial and ethnic differences between trial participants and non-participants. Costa et al. (2016) found that in multiple myeloma clinical trials between 2007 and 2014, 19.1% of people who participated were racial and ethnic minorities compared to 36.7%, the expected number of racial and ethnic minorities with multiple myeloma. A potential cause of this discrepancy is the current barriers to clinical trial accrual.

There has been extensive research conducted on barriers to clinical trial recruitment. Unger et al. (2016) distills the barriers into four main categories: "structural (especially, the absence of an available clinical trial), clinical (i.e., not meeting eligibility), attitudinal (with respect to both patients and physicians), and demographic and socioeconomic" (p. 187). While all barriers decrease clinical trial accrual, attitudinal barriers held by either HCPs or patients can be the most impactful (Unger et al., 2016). HCPs are in a unique position to be able to recommend treatment plans to patients. HCP preference for a specific treatment plan is one of the major factors influencing patient decisions (Jacobs et al., 2014; Eggly et al., 2008). However, while HCP preference is important, ultimately, it is the patient who decides to join a trial. It has been reported that some patients and their families hold perspectives that clinical trials present too many risks, and that they did not want their family member to be used as a "guinea pig" (Brown et al., 2013, p. 291). From a recent global survey, of the respondents worried about clinical trial safety, the majority were concerned about side-effects and had a general skepticism of the pharmaceutical industry as a whole (Anderson et al., 2018). A separate attitudinal barrier to clinical trial accrual is perceived cost. A 2007 survey of people with cancer revealed that 52% of patients were concerned with the cost of clinical trials (Meropol et al., 2007). This perceived cost is not unfounded. Though some assume clinical trials are free-of-cost, the American Society of Clinical Oncology's policy statement describes how through the various policies of both private and public healthcare plans, clinical trial participants can have high out-of-pocket costs due to an increased number of blood draws or scans, as well as the cost of the investigational tool in some cases (Winkfield et al. 2018). Attitudinal barriers about safety and cost could detrimentally impact clinical trial recruitment.

HCP and Patient Communication

An HCP is defined broadly as any professional who engages in care or treatment of patients or health conditions. Communication must be at the center of the HCP and patient relationship in order for the HCP and patient relationship to be successful (Ha et al., 2010). In the year 2018, 84.3% of adults and 93.6% of children in the United States came in contact with an HCP (U.S. Department of Health and Human Services, 2018a; U.S. Department of Health and Human Services, 2018b). Extrapolating from the percentages of people who interacted with an HCP in 2018, approximately 275 million people interacted with an HCP within 2018. As so much HCP and patient communication happens within a single year, it is important to further examine the harms and benefits of HCP communication.

Harms and Benefits of HCP Communication

Successful HCP and patient communication has numerous benefits (Sherwood et al., 2018). From a meta-analysis of HCP and patient communication research exploring patient adherence to care plans, it was found that successful communication is significantly correlated with increased patient adherence to care plans (Zolnierek & DiMatteo, 2009). It has also been found that successful communication leads to lower reported pain levels (Egbert et al., 1964). Egbert et al. (1964) conducted a survey in which they measured levels of narcotic usage after an electric intra-abdominal surgery. They found that patients who comprehensively explained the probable postoperative pain levels and strategies to alleviate their pain used fewer narcotics than control patients who were not informed of the potential pain (Egbert et al., 1964). Successful communication has

also been shown to lead to better patient outcomes (Lang, 2012; Stewart et al., 2000; Teutsch, 2003). Successful communication has also increased patient satisfaction (Lang, 2012; Haskard et al., 2009). HCPs who communicate more effectively have been found to increase the self-efficacy of HCPs (Nørgaard et al., 2012). Being a better communicator has also been found to improve the quality of the HCP and patient relationship (Rodriguez et al., 2008). On the contrary, poor communication unfavorably impacts patient management; poor communication is the leading reason why patients choose to bring forward a medical malpractice suit (Simpson et al., 1991; Avery, 1985).

Barriers to HCP Communication

After exploring the impacts of successful and unsuccessful HCP and patient communication, it is important to look at the barriers to successful HCP and patient communication. Two of the most significant communication barriers are language barriers and low patient health literacy (Tsoh et al., 2016). To illustrate, a survey of 1,200 Californians found that patients with limited English proficiency had a problem comprehending their visits with HCPs, difficulty comprehending the labels on medications, and a lack of understanding of the potential adverse reactions of medications or treatments (Wilson et al., 2005). Health literacy has also been associated with lower medication adherence in people with HIV, diabetes, and asthma (Osborn et al., 2007; Osborn et al., 2013; Soones et al., 2017).

Barriers to HCP and patient communication can be alleviated through the successful usage of good communication strategies. For example, patients experiencing chronic pain are more likely to be satisfied with their treatment if their HCP listened observantly, had sufficient time for the patient interaction, and made the patient feel as if the HCP was doing as much as they could to reach the patient's goal (Wiering et al., 2018). In a study within a population of people with prostate cancer, successful HCP communication was associated with how the HCP treated the patient within their interaction, how well the HCP knew the patient and his relevant medical history, and a greater sharing of information regarding the patient's diagnosis and prognosis (Song et al., 2014). A population of people with cancer also reported HCPs that were the most sensitive and empathetic were the most welcomed by the patients, increasing trust and rapport (Bridge et al., 2019).

Patient-Centered Communication

All of these communication strategies associated with positive patient outcomes can fall under the umbrella terms of patient-centered communication. Patient-centered communication was first defined by the Institute of Medicine in 2001 as "providing care that is respectful of and responsive to individual patient preferences, needs, and values, and ensuring that patient values guide all clinical decisions" (pg. 3). Shared decision making has been described as the "pinnacle of patient-centered care" by Michael J. Barry and Susan Edgman-Levitan in the title of their 2012 article in the New England Journal of Medicine (pg. 780). Though there are some medical cases in which there is one clear step forward, for example, surgery to repair a broken limb or sutures to close a large wound, the majority of medical cases require a decision by either the HCP or patient to move forward (Barry & Edgman-Levitan, 2012). It is in this context that shared decision making becomes significant (Barry & Edgman-Levitan, 2012). For shared decision making to occur, the HCP must provide information to the patient about potential treatment options and the risks and benefits to each treatment (Barry & Edgman-Levitan, 2012). The patient must then share their desires and beliefs about the treatments (Barry & Edgman-Levitan, 2012). With both parties now equally informed, shared decision making can occur as both parties agree with what would be the best treatment. If fictional media is representing empathetic, patient-centered care, this could be a positive influence on potential clinical trial recruits.

Clinical Trial Communication

In order to better contextualize the analysis of fictional films and television episodes, it is important to gain a fuller understanding of existing literature surrounding health practitioner and patient communication in the context of clinical trial recruitment. Albrecht et al. (1999) conducted a

groundbreaking study in which they visually recorded conversations between HCPs and people with cancer. The researchers sought to distinguish specific techniques of communication that culminated in patient accrual to the clinical trial. They found that patients who were recruited to clinical trials had HCPs who more often described the clinical trial as research and mentioned the benefits of the clinical trial. They found there was not a difference in the number of apprehensions that recruited patients versus unrecruited patients had, nor in the time that the patients talked to the HCPs. Albrecht et al. (1999) specifically found that patients in oncology clinical trials were more likely to consent to medical research when they were told the potential benefits and were verbally given the information within the informed consent documents instead of only being presented a paper document.

From Albrecht's first work until today, several different aspects of clinical trial communication have been analyzed. One aspect is whether or not an HCP recommends a patient to join a clinical trial. Susan Eggly and her research team (2008) analyzed oncologist recommendations for patients and how they impact recruitment to clinical trials. They looked at whether an oncologist did or did not provide a recommendation, the context that the recommendation was given in, and the type of recommendation. They found that in 68% of the patient-HCP exchanges, at least one positive recommendation was given to begin a trial. Over three-fourths (77%) of the positive recommendations were unsolicited, while the remainder (23%) of recommendations came after a patient directly asked (Eggly et al., 2008). Eggly et al. (2008) found that 77% of the recommendations were directed toward the patient explicitly, explaining that the trial would have direct benefits to the patient, while the other 23% of recommendations pointed to a positive quality of the trial itself. They found a significant positive relationship between a healthcare provider's recommendation and a patient's decision to participate in a trial. Eggly et al.'s (2008) study is informative as it demonstrates the patient's susceptibility to an HCP recommendation and the information-seeking behavior of patients.

Due to the adaptive and flexible nature of communication, it is impossible to explicitly list specific behaviors that lead to successful communication. However, communication that is patient-centered and empathetic is the ideal. Clearly, successful HCP communication has various implications, and this holds true in the context of clinical trial decision making. The ways that patients learn to interact with HCPs are diverse. Patients receive messages about HCPs and the health industry in numerous ways. Some messages are received during direct contact with HCPs, such as within a clinic visit or a visit to the emergency room. Still, other messages are passively consumed by viewing mass media representations of healthcare.

Mass Media Representations of Healthcare

Compston (2006) describes mass media as a media form intended to spread a message to a large number of people. He defines mass media as made up of both electronic and audiovisual media and print media. Electronic and audiovisual media is made of "radio and television… movies, pre-recorded videos and DVDs, sound recordings…, and new media such as the Internet and video games" (Compston, 2006, pg. 217). Mass print media is made up of "books, journals, magazines, and newspapers" (Compston, 2006, pg. 217).

Mass media can depict a vast array of messages. For this thesis, the mass media representations of science, specifically healthcare, will be discussed. The mass media representations of science and healthcare impact viewers of the media (Rideout, 2008; Dudo et al., 2011; Hether et al., 2008; Quick, 2009). However, those impacts can be variable. Studies have been completed to measure the direct impacts of specific messages in mass media. Hether et al. (2008) compared knowledge and attitudes toward BRCA1 mutations and their impact on breast and ovarian cancer incidence rates. The researchers randomly surveyed 559 female primetime television viewers before and after watching a two-episode arc of ER or a single episode of Grey's Anatomy in which two female patients dealt with surgery decision-making after learning they had a BRCA1 mutation. They found that there was an additive effect

between the episodes, with more favorable views of surgical treatment plans for BRCA1 breast cancer when the respondents had seen all three episodes. This study depicts the positive effects that certain storylines may have on viewers. In a similar study, Rideout (2008) details the impact that a crafted storyline on Grey's Anatomy had on public understanding of the vertical transmission of HIV from mother to baby. In this study, The Kaiser Family Foundation, a firm focused on health messaging, worked with Grey's Anatomy screenwriters to incorporate a storyline that could educate watchers on a health topic. Rideout (2008) first surveyed known Grey's Anatomy watchers one week before the episode aired, then both one week and six weeks after the episode aired. Compared to surveys taken before the episode aired, Rideout (2008) found that respondents had significantly higher correctly reported rates of vertical HIV transmission with treatment both one week and six weeks after the episode aired. Though the six-week responses were reduced in accuracy, they still were higher than the original baseline rates (Rideout, 2008). This study illustrates how crafted storylines can have specific impacts on the viewer's knowledge. Finally, studies have also been conducted to specifically measure behavior following an interaction with a mass media message. For instance, Quick (2009) conducted a survey of college students to measure perceptions and willingness to participate in organ donation after watching the ten-episode arc of heart transplant recipient Denny Duquette on the television program Grey's Anatomy. The survey compared non-watchers to loyal watchers to assess two myths of organ donation: that wealthy people can use their wealth and power to advance on the transplant list— "the purchase myth"— and that knowing specific HCPs can increase a person's chance to get a donor organ—"the relationship myth" (Quick, 2009, pg 788). The survey found that loyal viewers were less likely to believe the purchase myth, but there was not a significant difference between surveyed groups in terms of the relationship myth (Quick, 2009). On the other hand, loyal viewers were more likely to have responded that they would discuss organ donation with their family members, but no differences in intentions to donate organs in the future were seen (Quick, 2009). In sum, these studies suggest that mass media can have real and measurable impacts on viewer attitudes and behavior.

Mass Media Representations of HCP and Patient Clinical Trial Communication

Media representations of science and technology can have an impact on public perception and attitudes, which can lead to altered actions as exemplified through Hether et al. (2008), Rideout (2008), Quick (2009), and others. These studies show that fictional media can be used as a tool to influence consumers of fictional media. To explore the substance within media depicting clinical trials, Fisher and Cottingham (2017) conducted a content analysis of the themes of films and television episodes that have storylines detailing medical research. They sampled 65 television shows and films depicting medical research and analyzed the recruits' "race, gender, class, and motive of the participants represented; and the depiction of informed consent for and the outcome of the research" (p. 567). They also used qualitative media analysis to analyze key themes in fictional media. The majority of the fictional research participants were white males, a fact that mirrors current clinical trial participation (Fisher and Cottingham, 2017; U.S. Food and Drug Administration, 2017). Fisher and Cottingham thought it was problematic that the portrayals of medical research had mostly negative outcomes for the patients and that clinical trial recruits were majority white and male. Though these findings are still important overall, it is essential to explore the HCP communication within these fictional clinical trial scenarios to understand barriers to clinical trial recruitment better. Specifically, it is important to examine HCP and patient communication as HCP interactions are the driving force shaping patients' perceptions of HCPs. Though Fisher and Cottingham (2017) have completed an overall content analysis of the fictional representations of clinical trials in television and film, a content analysis of HCP and patient communication surrounding clinical trials has not been completed.

In sum, the key questions that arise from the literature are as follows:

1. What is the quality of HCP to patient communication surrounding clinical trial scenarios in fictional media?
2. Do depictions of HCP and patient communication in fictional media with clinical trial scenarios alter perceptions of clinical trials and potential to join a trial in the future? and
3. Can the content in fictional media portraying HCP and patient communication regarding clinical trials be altered to increase recruitment to clinical trials?

To address question one, I aim to analyze television interactions between HCPs and patients through a content analysis in which both quantitative and qualitative communication features are measured. Through this analysis, the features of fictional clinical communication in television will be more accurately described. My findings will motivate future research to more accurately measure the impact of fictional clinical trial communication on the future potential to join a clinical trial and to explore whether fictional media content can be altered to increase clinical trial recruitment- addressing the second and third questions called for by the literature. My findings will also better inform HCPs about possible pre-held patient beliefs regarding clinical trials.

METHODS

Criteria

A structured content analysis was completed to analyze the portrayals of fictional HCP and patient communication. Seventy-nine films and television episodes ranging from 2004 to 2018 were initially screened (Appendix A). Sixty-five television episodes and films came from Fisher and Cottingham's 2017 content analysis that aired between 2004 and 2014. Fourteen additional television episodes familiar to the author were also included. All 14 episodes were aired in 2018, after the publication of Fisher and Cottingham (2017). No episodes were analyzed from before Fisher and Cottingham's 2017 content analysis as I wanted to explore current media portrayals. Television episodes were exclusively analyzed due to ease of access and time restraints for analysis. Thus, from the original 79 media representations, 60 television episodes were included, and 19 films were excluded (Figure 1).

Of those 60 television episodes, 38 were available on streaming services Netflix, Hulu, or Amazon Prime Video. All 38 of those episodes were included and screened for HCP and patient interactions surrounding clinical trials, while the remaining 22 episodes were excluded (Figure 1). The 38 episodes were screened to ensure there was at least one conversation between an HCP and a patient, family member, or friend. The time points that each interaction began and ended were recorded. The general description of the interaction and overall outcome for the patient were also documented. Of the screened episodes, only 23 episodes with 90 individual communication events were included (Figure 1). The remaining 15 episodes were excluded because they either did not depict an HCP talking to a patient about a clinical trial or were not medical dramas. Only medical dramas were included as they had more realistic interactions that could potentially shape the viewer's behavior more considerably.

After the original screening, there were a total of 90 individual communication events. All episodes were then screened again. At this point, if each interaction was significant, it was included and analyzed qualitatively and quantitatively. Interactions were deemed insignificant and excluded if the interaction was less than one sentence or less than five seconds, if an HCP was not present within the conversation, or if the clinical trial or experimental treatment was not referred to explicitly. Communication events between an HCP and a patient, an HCP and a patient and their family, and an HCP and the patient's family were included. Of the 90 individual communication events analyzed, 57 interactions from 15 television episodes were ultimately included in the sample (Figure 1, Table 1).

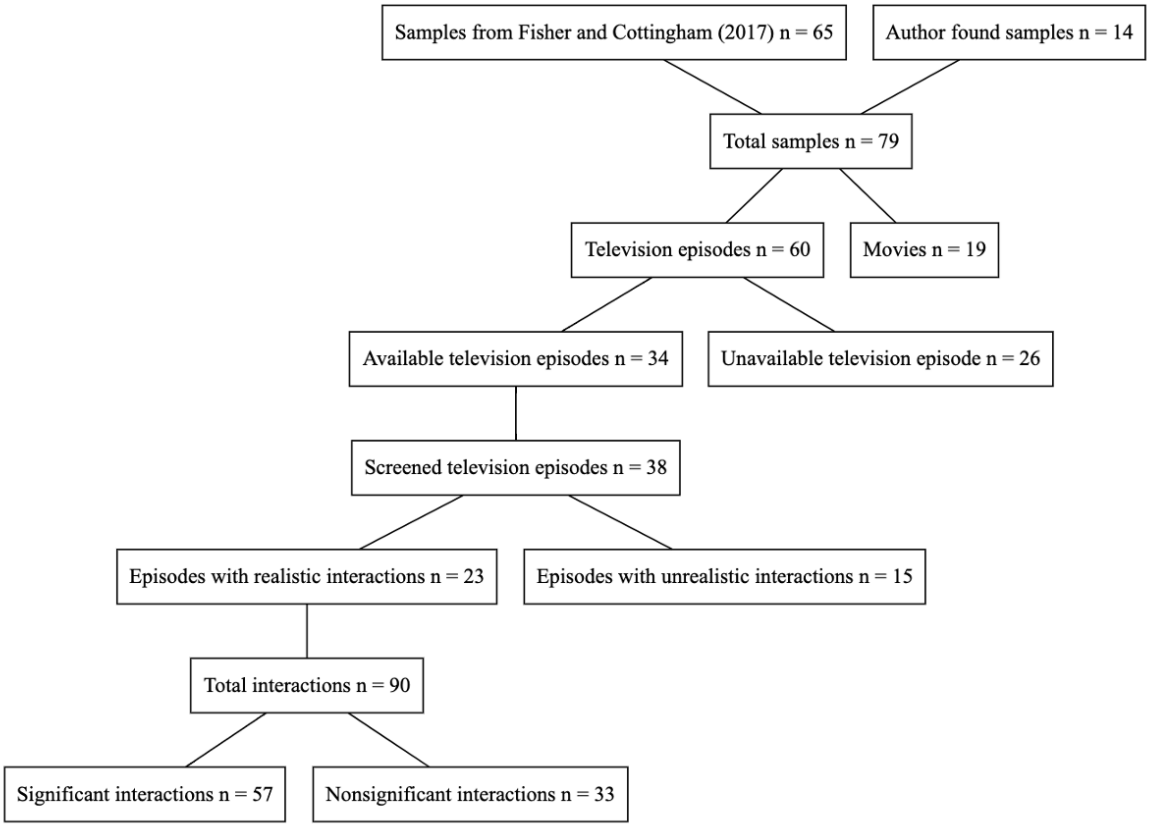


Figure 1: Inclusion and exclusion of episodes and communicative interactions.

Analysis

Quantifying HCP Communication

The 57 communication events were quantified using several methods. To determine whether HCPs or patients spoke a larger proportion of the time, the number of seconds that each person was speaking was measured. Each individual HCP's speaking time and individual patient's speaking time during an interaction was measured. All family members' or friends' speaking times were combined as there were only two episodes with multiple family members or friends within the room. In that case, all members had the same goals and were not in conflict so that they could be analyzed as one unit. To more easily compare HCP speaking time to patient, family, and friend (PFF) speaking time, all individual HCP speaking times were totaled regardless of the number of HCPs speaking. In some cases, only family members or friends spoke during the interactions with an HCP, so all PFF speaking times were totaled as well. A ratio was then computed of total HCP speaking time to total PFF speaking time. These ratios are the speaking scores for each interaction. An interaction with an HCP speaking for a greater proportion of the conversation has a score of greater than one, while an interaction with PFFs speaking for a greater proportion of the conversation has a score of less than one.

To determine whether HCPs or PFFs asked a larger proportion of questions, the number of questions that either person asked was also counted. A sentence or phrase was considered a question if it was phrased as a traditional question beginning with "who," "what," "when," "where," "why," or "how;" if within the subtitles there was a question mark at the end of the phrase; or if the inflection the phrase indicated it was a question. To more easily compare HCP's number of questions to PFF's number of questions, all HCP questions were grouped and totaled per interaction, and all PFF questions were grouped into and totaled per interaction. As with the speaking times, a ratio of HCP to PFF questions

was computed to create scores to identify whether HCPs or PFFs were asking a greater number of questions. In some interactions, either HCPs, PFFs, or both asked zero questions. Because of these zero counts, when calculating the ratio scores, undefined and infinite scores were seen. To account for this, the totaled number of questions spoken by HCPs and PFFs were transformed by adding 0.5. The transformation made all of the calculated scores finite and real. This is similar to the Haldane-Anscombe correction for calculating odds-ratios (Haldane, 1940; Anscombe, 1956). As with the speaking scores, scores over one represent HCPs asking a greater proportion of questions, and scores below one represent patients asking a greater proportion of questions.

Measuring Quality of HCP Communication

Huntley et al. (2012) created the Liverpool Undergraduate Communication Assessment Scale (LUCAS) to measure HCP communication from the framework that communication should primarily meet the patient's needs. Taking into account the way communication is by its nature organic, creative, and dependent on context, the LUCAS evaluates the way an HCP meets the patient's goals within an interaction rather than the HCP doing specific communicative tasks (Huntley et al., 2012) (Appendix B).

To better measure televised fictional HCP and patient communication specifically, I adapted LUCAS. I removed the "A) Greeting & Introduction" and "B) Identity check" portions as in many fictional representations, these steps had been skipped because the HCPs already had an established relationship. I also removed the "C) Audibility and clarity of speech" section. The nature of television requires that all speakers be heard, so this section was not relevant for my analysis. The included sections were "Non-verbal behavior" measuring the non-verbal behavior of the HCPs, "Questions, prompts, and/or explanations" measuring the way an HCP asked and answered questions, prompted information from the patient, and explained procedures and treatments, and "Empathy and responsiveness" measuring whether the HCP took the patient's needs, views, and feelings into account during the interaction (Appendix C). In each section, the HCP could score zero for a poor performance, one for a moderate performance, and two for an excellent performance (Appendix C). Overall, the lowest total score an interaction could receive was zero, and the highest was six. These sections closely correlate with ideal patient-centered communication, making the adapted LUCAS a useful tool for measuring communication quality. The LUCAS scores were used as an overall categorical measure of the communication interactions. Interactions scoring between zero and two were negative communication interactions. Interactions scoring three or four were neutral interactions. Interactions scoring five or six were positive communication interactions.

Results

To explore the quality of HCP communication in fictional representations of HCP and patient interactions regarding clinical trials, communication events in 15 episodes of television were quantified and the quality was measured. The episodes ultimately included were from television shows Grey's Anatomy, House M.D., and All Saints and aired between 2008 and 2018 (Table 1). In terms of overall HCP communication quality, a statistically significant number of interactions were positive (p-value < 0.05). Over two-thirds, 84.2% (n = 48/57), of interactions were positive in nature, while a small percentage were neutral and negative, 7.0% (n = 4/57) and 8.8% (n = 5/57), respectively (Figure 2).

To investigate the outcomes of patients in the context of the storyline and clinical trial best practices, the patient's outcome was compared to the portrayal. Overall, more patients had negative outcomes (n = 8) than positive outcomes (n = 4) (Figure 3). Negative outcomes are outcomes which the PFF did not desire. In the negative outcomes, two patients received placebos after hoping to receive the investigation agent, two healthy volunteers had adverse drug reactions, and four patients died (Figure 3). In the positive outcomes, three patients received the desirable active agent, and one patient had reduced tumor size (Figure 3).

To explore whether HCPs or PFFs spoke a greater proportion of time per interaction, a ratio of total HCP to total PFF speaking time was calculated. Both negative and positive ratings had average ratios in which HCPs spoke a greater proportion of time than PFFs, 2.15 and 2.66, respectively (Figure 4). Neutral interactions had an average ratio of 0.28 in which PFFs spoke greater amounts of time per interaction (Figure 4). There was no clear correlation between the overall communication rating and the ratio of speaking times. On average, the number of questions asked by HCPs was lower than the number of questions asked by PFFs in both negative and neutral interactions, with a ratio of 0.82 and 0.6 respectively (Figure 5). In positive interaction, HCPs asked more questions than PFFs with a ratio of 1.68 (Figure 5).

Title	Season	Episode	Year	Trial context
Grey's Anatomy	4	13	2008	Brain tumor surgery trial
Grey's Anatomy	4	14	2008	Brain tumor surgery trial
Grey's Anatomy	4	15	2008	Brain tumor surgery trial
House M.D.	5	3	2008	Healthy volunteer in multiple drug trials
House M.D.	5	14	2008	Huntington's drug trial
All Saint's	12	6	2009	Healthy volunteer in drug trial
Grey's Anatomy	7	13	2011	Alzheimer's treatment trial
Grey's Anatomy	7	16	2011	Alzheimer's treatment trial
Grey's Anatomy	7	17	2011	Alzheimer's treatment trial
Grey's Anatomy	7	19	2011	Alzheimer's treatment trial
Grey's Anatomy	7	22	2011	Alzheimer's treatment trial
Grey's Anatomy	15	22	2018	Quadriplegic in stem cell trial
Grey's Anatomy	15	23	2018	Quadriplegic in stem cell trial
Grey's Anatomy	15	24	2018	Quadriplegic in stem cell trial
Grey's Anatomy	15	25	2018	Quadriplegic in stem cell trial

Table 1: Sampled television episodes by year

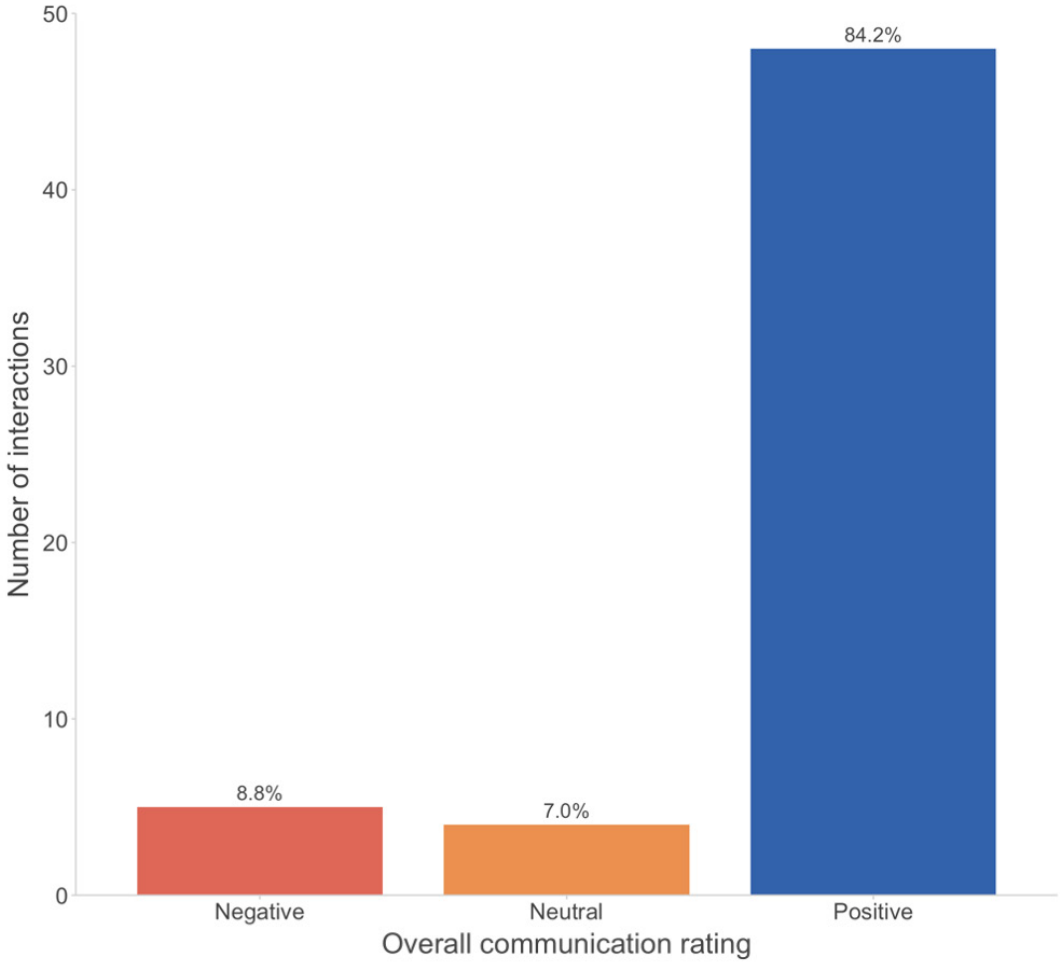


Figure 2: A vast majority of HCP communication events are categorically positive

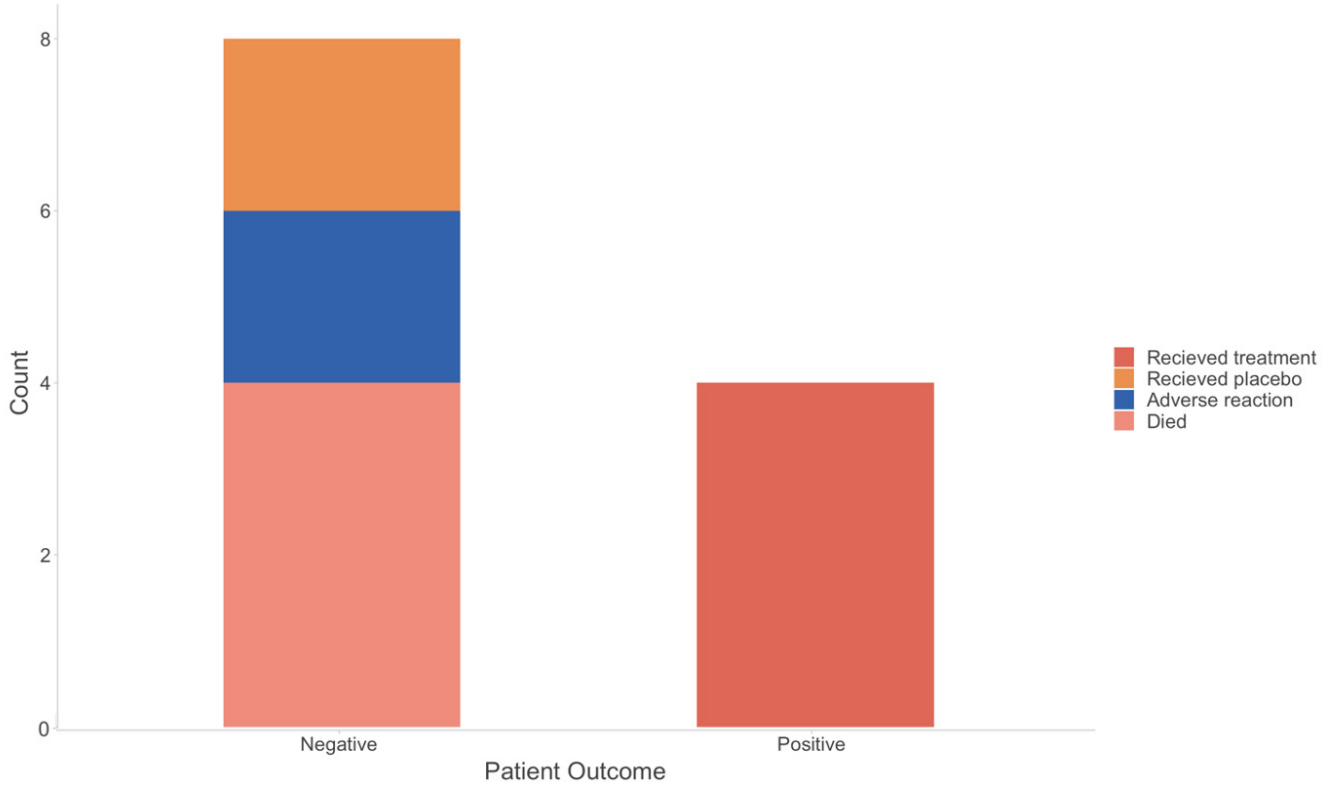


Figure 3: More patients were portrayed as having negative outcomes than positive outcomes

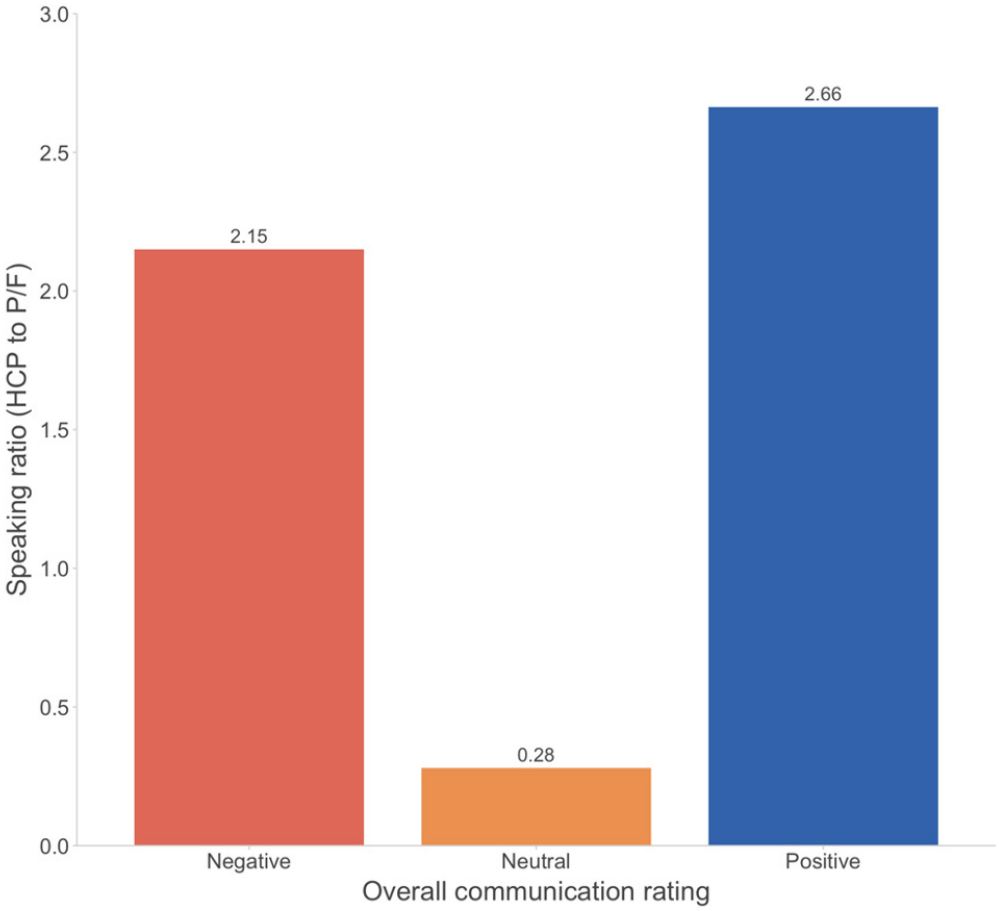


Figure 4: In negative and positive communication events, HCPs speak twice as long as PFFs

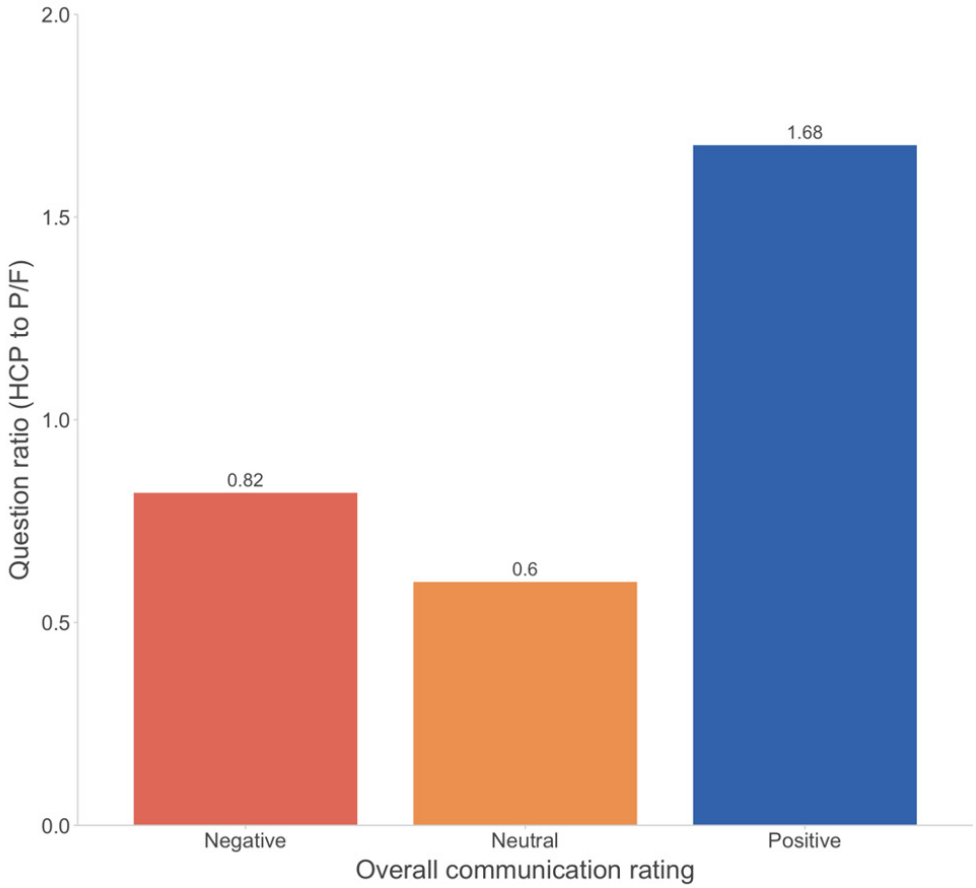


Figure 5: HCPs ask more questions than PFFs in positive communication events

Discussion

To characterize the HCP communication found with televised representation of clinical trials, I analyzed the communication with 15 television episodes. More than two-thirds of the communication events analyzed were categorically positive, scoring highly on the adapted LUCAS scoring system (Figure 2). A majority of the patients had negatively framed outcomes, though the context of the interaction is important to discuss (Figure 3). In terms of the speaking and question ratios, there was not any clear association between the quality of the communication and whether HCPs or PFFs spoke longer or asked more questions (Figure 4). Overall, HCPs utilized patient-centered communication and were empathetic and respectful regardless of the context of the interaction.

My findings demonstrate that in televised fictional representations of clinical trial communication, HCP communication was productive and successful between HCPs and PFFs overall. In the 15 episodes included in the analysis, four patients died as a result of an experimental treatment, two patients received placebos which were framed negatively in the episode, and two healthy volunteers had adverse reactions to novel drug treatments. However, when taking into account the medical context, these seemingly negative portrayals appear more neutral or positive in nature.

A first-in-man experimental treatment using an oncolytic virus on non-operable brain tumors was depicted in all analyzed episodes in season 4 of Grey's Anatomy. All HPCs communicated clearly to all patients and PFFs that this surgery was one with multiple risks and was a last-ditch effort to reduce their tumor size. To illustrate, when talking with Phillip, a patient with a brain tumor, before the surgery, Dr. Shepard said, "Phillip, this is an extremely risky operation. You still may have a few weeks to live. You could spend that time with Jennifer. You don't have to do the surgery today" (McKee & Tinker, 2008). On a separate occasion, in response to a Phillip stating, "Well, if the surgery doesn't kill me, the tumor will," Dr. Shepard responded, "Well as we discussed, the treatment has never been tested on humans before" (McKee & Tinker, 2008). These quotes exemplify the way that HCPs are depicted as communicating the known and unknown risks of a first-in-man clinical trial. Portraying risks is a positive trait of fictional HCP communication in the context of clinical trials. Portraying risks accurately allows viewers to witness how the risks in clinical trials would be communicated if the viewer was participating in a clinical trial.

In total, HCPs were shown communicating with five patients before treatment. Four of the patients died during the treatment, while the fifth and final patient survived and was shown to have decreased tumor size. The fact that 80% of patients in the episodes did not survive the treatment may seem dismal when not looking at the medical context. However, taking into account that the patients were only weeks away from death lessens the impact of the deaths. Patients themselves spoke phrases like, "Well, if the surgery doesn't kill me, the tumor will. What have I got to lose, right?", "Hey, at least we don't have to worry about [pain], you know? The surgery will work, okay? And if it doesn't at least we'll die quick", and "I'm just saying, if this works, my name will go down in the annals of some book. That's kind of cool" (McKee & Tinker, 2008; Rhimes & Corn, 2008). In these examples specifically, the medical context provides a small justification for the deaths. Though death is generally seen as a negative, patients going into surgery are given hope, juxtaposing the potential of death with the possibility of success when all other hope of survival is gone.

Season 7 of Grey's Anatomy portrayed a clinical trial for people with Alzheimer's disease to potentially slow down or reverse the disease progression. The episodes depicted two patients who received placebo treatments, while two patients received the active agents. In these episodes, the families of patients often urged the administrators of the trial to give their family member the active agent. Again, in this context, the patients and their families were willing to go to vast lengths to stop the disease progression. For example, in this scenario, a patient's wife, Mrs. Cobb, begs Dr. Alex Karev to give her husband the experimental drug even if he is slated to receive the placebo.

MRS. COBB: Please. Give my husband the drug.

DR. KAREV: Mrs. Cobb…

MRS. COBB: You could do it, if you wanted to. You know that you could.

DR. KAREV: No, I couldn't, it's completely random. The computer decides. We don't even know if he's getting the drug until we're about to inject it.

MRS. COBB: Well then change it if you need to. I need my husband back; I need Daniel back.

DR. KAREV: I'm sorry, I just... even if the drug worked. I'm not going to reverse things.

MRS. COBB: But I can live with that, I- I can. I can live with what-what-what- we have right now. I can live with being a part-time wife. But I cannot live with losing my husband all the way. So... for god sake, give him the drug. Please (Wilding & McKidd, 2011).

In a second example, Dr. Richard Weber, who is the chief of surgery of the hospital, attempts to convince Dr. Derek Shepard to put Dr. Weber's wife Adele on his Alzheimer's trial.

DR. WEBER: There's always a way to work the system, Derek. We both know that.

DR. SHEPARD: I can't put Adele in my trial, Richard. I've looked at it all day.

DR. WEBER: Well, go over it again.

DR. SHEPARD: You're not hearing me. I love Adele, but I can't pull a patient off the waiting list. It could compromise the whole trial. It could get us blacklisted by the FDA.

DR. WEBER: Look, Adele has no future without this. (Nowalk & Ornelas, 2011).

In both of these examples, the family members of patients are asking HCPs to break the rules of the clinical trial. In the second example, a member of leadership in the hospital is the one putting the pressure on an HCP that he supervises. These examples show the desperation that family members feel to help their ill family member, but they also show the resilience of HCPs to not deviate from the proper policies of the clinical trial. These examples demonstrate that negative patient outcomes do not inherently depict clinical trial negatively; rather, the negative patient outcome can be both a function of the plot and also act to depict realistic occurrences within clinical trials.

In the episodes depicting patients receiving placebos, the communication about placebos had a more negative tone. For example, after a patient joked about receiving the "wonder drug," Dr. Shepard responded, "Well, if it came down to how much I like the patient, you'd be a shoo-in" (Wilding & McKidd, 2011). In this example, the HCP implied that he wished that patients he liked received the experimental drug and not the placebo. Receiving a placebo treatment is a primary fear amongst patients when they are considering whether or not to opt into a clinical trial (Meropol et al., 2007; Stone et al., 2005). Language like this could further attitudinal barriers people have against joining clinical trials. Nevertheless, it is still important to depict patients receiving placebo treatment as placebo-controlled clinical trials frequently occur, especially when there is not a standard-of-care treatment available. While overall, HCPs respected policies about the blind placebo-controlled trial, there was an instance in which the policies were violated. Dr. Meredith Grey, a resident at the hospital, first broke policy by looking into whether a patient who was a family friend would receive the treatment or a placebo. Dr. Grey then changed what treatment the patient would get, from placebo to active treatment. This was an ethical violation on the part of the HCP. She should not have changed the treatment as it jeopardized the partiality of the entire clinical trial. In the storyline of this episode, the HCP received punishment as she was fired. This helps to show the viewer that this behavior is not warranted nor acceptable in a real clinical setting.

After examining the overall communication trends in the various interactions per episode, the proportion of HCP to PFF speaking times and the proportion of HCP to PFF questions were examined.

No clear trend between the overall communication rating and either the speaking ratio or question ratio was seen (Figure 5, Figure 6). However, in both negative and positive communication events, HCPs spoke at least twice as long as PFFs (Figure 5). In neutral speaking events, PFFs spoke longer than HCPs. These results are partially expected. In most medical contexts, HCPs typically speak a greater amount of time during a visit than patients (Peck and Conner, 2011; Hagiwara et al., 2013; Street et al., 2014). These results are also consistent with the LUCAS rating scale. LUCAS does not take into account quantitative measures like speaking time or numbers of questions, so a correlation between speaking and question ratios is not inherent to the qualitative ratings, unlike other communication measurements. This also demonstrates that the writers of medical dramas do not use a clear formula when scripting communication events. The written communication is intrinsically flexible and adaptable to each interaction, and the specific goals of characters are what determine the communication choices.

In contrast, Fisher and Cottingham (2017) found that the overall representations of fictional clinical trials were negative, with poor patient outcomes. Since Fisher and Cottingham (2017) analyzed the representations from a different perspective than the one in this paper, it is understandable how they conclude that the representations were overall negative. Though a majority of fictional patients had negatively framed outcomes in the episode I analyzed, when taking into account the context of the interaction, the representations analyzed were positive. There are several reasons why Fisher and Cottingham (2017) and I drew different conclusions. Primarily, I was analyzing HCP and PFF communication in medical dramas only, while Fisher and Cottingham (2017) explored the general patient outcomes in various genres such as comedies and science fiction. Communication in medical dramas was more prototypical of genuine interactions with HCPs than communication in other dramas. Because the communication is more similar to real communication, it was more likely to be categorically positive than the overall interaction that Fisher and Cottingham (2017) explored. Also, four episodes that aired after Fisher and Cottingham (2017) was published were included in my analysis. In these episodes, the patient depicted had a positive outcome. This would not have been in Fisher and Cottingham's (2017) analysis. This might also point to a positive trend in clinical trial representations, with more recent episodes being increasingly positive.

Cottingham and Fisher (2017) followed up their initial content analysis with interviews of healthy volunteers of clinical trials. The volunteers would spontaneously bring up negative media portrayals of clinical trials (Cottingham and Fisher, 2017). Cottingham and Fisher (2017) argue that healthy volunteer research participants used fictional representations to "manage their anxieties surrounding their involvement in research" (p. 292). Using humor, the volunteers would refer to what they would describe as outlandish symptoms or side effects of medication within the trials. The humor can be seen as a coping strategy to handle trial-related anxieties. For example, an interviewee stated, "That's why people's perceptions are so negative of clinical studies 'cause they start thinking like horrible experiments gone wrong, with-with Hollywood" (p. 295). Statements like these led Cottingham and Fisher (2017) to the conclusion that "there may be individuals who are dissuaded from participating in clinical trials because of the ways in which scientific research is portrayed in popular culture" (p. 297). Further, they state that the friends and family members of the healthy volunteers were quoted saying that they would not join a clinical trial due to fictional representations (Fisher and Cottingham, 2017). This article shows that fictional representations can have a clear impact on clinical trial recruitment with healthy volunteers and their friends and family. From the episodes I analyzed, I understand how healthy volunteers could be apprehensive about clinical trial involvement. In the two episodes portraying healthy volunteers, both had adverse reactions to experimental treatments. Both healthy volunteers ultimately recovered and appeared not to have long term health issues, but one healthy volunteer was spoken to negatively as he did not follow the trial's guidelines of only participating in one trial at a time. Overall, healthy volunteers were represented more poorly than people who were ill and participating in trials. The poor representation could be because healthy volunteers were represented as being motivated financially; while people

who were sick, were represented as being motivated to heal themselves or to further science. If media professionals cooperate with television writers, fictional television representation of clinical trial communication could be tailored to aid in recruitment. If more positive communicative interactions are shown, then recruitment could be increased, and healthy volunteers would not rely on negative portrayals to curtail their anxieties about clinical trial participation.

Limitations

This study has several limitations. In total, only 14 episodes were included in the analysis, and all were medical dramas. The episodes aired in the years 2008, 2009, 2011, and 2018. These are modern media sources and may not represent the totality of fictional clinical trial representations from earlier periods. Also, the low number of episodes, coupled with the lack of variety in genre and the non-normal distribution in years that the episodes aired, led to results that cannot be widely generalizable to other genres. The gaps in the year that episodes aired made it unrealistic to examine relationships between the overall communication rating and time. If more episodes from a greater variety of years and genres were included, the results would be more generalizable, and the communication trends over time could become evident.

Future Directions

My findings are an essential first step in analyzing the impact that fictional health communication can have on clinical trial communication. To further explore these concepts, there are several proposed experiments. First, while my work examines modern medical drama representations of fictional HCP and patient communication in the context of clinical trials, it is vital to expand the genres and years of future analyzed work. Interesting findings could result from analyzing more non-realistic portrayals, such as in comedies or science fiction. If the non-realistic portrayals are compared to the realistic portrayals, there should be differences in content and communication. Even from the initial screening of episodes, these differences were evident though out of the scope of the thesis. These findings could also be incorporated into similar studies, as proposed below.

While an understanding of the content of HCP and PFF communication within the clinical trial context in fictional television episodes is beneficial, it is more important to understand the impact that the portrayals can have on the viewers' future potential to join clinical trials. I propose several future studies similar to Hether et al. (2008), Quick (2009), and Rideout (2008). As discussed previously, Quick (2009) surveyed college-aged students to determine if viewing an organ donation storyline on Grey's Anatomy impacted their willingness to donate organs in the future or talk with friends and family about organ donation. A similar multi-phase study would be useful to measure whether viewers' perceptions of clinical trials are vulnerable to changing after watching fictional HCP communication in the context of clinical trials. For phase 1, the study would be conducted via random survey-takers. The participants would first complete a survey that measures their attitudes toward clinical trials and willingness to participate in future clinical trials. The participants would then be randomly assigned to watch a video clip of the positive, neutral, or negative communication interactions per this content analysis. The participants would immediately be surveyed after watching the clip, and then six weeks after. Multiple surveys allow for a better understanding of whether the fictional clips of HCP communication made an immediate or sustained impact on willingness to participate in clinical trials and attitudes toward clinical trials. The second phase would consist of asking television writers to create storylines about clinical trial recruitment or the conducting of a clinical trial. Similar to Rideout (2009) and Hether et al. (2008), I propose working with television and film writers within various genres to create story arcs to educate the public on clinical trials. The storylines could be adapted to work within the context of each genre. They should be accurate and informative about how people are recruited to clinical trials, as well as how clinical trials are conducted. Hopefully, several short films over various genres could be created, and surveyed participants could come and view them. As before, survey-takers would be surveyed about willingness to participate in clinical trials as well as perceptions of clinical trials before, immediately after, and six

weeks after watching the short films. If the short films are effective educators of the public, then this strategy of working with media professionals to advance knowledge of clinical trials could be an effective strategy to increase clinical trial participation.

Conclusion

As the costs for research and development for novel pharmaceuticals continues to surge, it drives the out-of-pocket costs for prescription medications to skyrocket. Thus, it is crucial to explore all potential reasons for increasing costs. A clinical trial's failure due to low recruitment causes an extreme and disheartening expense of pharmaceutical research and development. The lost expense of the clinical trial can then be translated directly into higher drug costs. When a trial fails due to a lack of drug safety or efficacy, researchers gain negative results from that experiment. However, when a clinical trial fails due to low recruitment, this is a total loss of data. Neither positive nor negative results can be extrapolated if the power of the study is low due to low trial recruitment. Thus, barriers to clinical trial recruitment must be examined.

Overall, I found that HCP communication surrounding clinical trials in medical dramas is patient-centered, empathetic, and categorically positive in nature. These results counter the claims made by Fisher and Cottingham (2017), who stated that clinical trials are generally depicted negatively due to the outcomes of people in the clinical trials. This difference is essential as contextualized communication can be most telling. Taking into account the context and the communication within the interaction establishes justification in terms of the preservation of life and advancement of science. This study illustrates an opportunity for health communicators working with the media community to better tailor fictional communication and improve messaging about clinical trials. The improved messaging could educate more people about how real clinical trials function, hopefully leading to greater participation and lessened fear. The increased participation could help decrease overall pharmaceutical costs as trials will fail due to low recruitment less often and potentially improve clinical trial recruitment.

References

Albrecht, T. L., Blanchard, C., Ruckdeschel, J. C., Coovert, M., & Strongbow, R. (1999). Strategic physician communication and oncology clinical trials. *Journal of Clinical Oncology*, 17(10), 3324-3332

Anderson, A., Borfitz, D., & Getz, K. (2018). Global public attitudes about clinical research and patient experiences with clinical trials. *JAMA Network Open*, 1(6), e182969. doi: 10.1001/jamanetworkopen.2018.2969

Anscombe, F. J. (1956). On estimating binomial response relations. *Biometrika*, 43(3/4), 461-464. doi: 10.2307/2332926

Attarwala, H. (2010). TGN1412: from discovery to disaster. *Journal of Young Pharmacists,* 2(3), 332-336. doi: 10.4103/0975-1483.66810

Avery, J.K. (1985). Lawyers tell what turns some patients litigious. *Medical Malpractice Review*, 2, 35-37

Barry, M. J. & Edgman-Levitan, S. (2012). Shared decision making—the pinnacle of patient-centered care. *The New England Journal of Medicine,* 366, 780-781. doi: 10.1056/NEJMp1109283

Braunholtz, D. A., Edwards, S. J. L., & Lilford, R. J. (2001). Are randomized clinical trials good for us (in the short term)? Evidence for a "trial effect". Journal of Clinical Epidemiology, 54(3), 217-224. doi: 10.1016/S0895-4356(00)00305-X

Bridge, E., Conn, L. G., Dhanju, S., Singh, S., & Moody, L. (2019). The patient experience of ambulatory cancer treatment: A descriptive study. *Current Oncology*, 26(4), e482-493. doi: 10.3747/co.26.4191

Brown, R. F., Cadet, D. L., Houlihan, R. H., Thomson, M. D., Pratt, E. C., Sullivan, A., & Siminoff, L. A. (2013). Perceptions of participation in a phase I, II, or III clinical trial among African American patients with cancer: What do refusers say?. *Journal of Oncology Practice*, 9(6), 287-293. doi: 10.1200/JOP.2013.001039

Buonansegna, E., Salomo, S., Maier, A.M. & Li-Ying, J. (2014), Why do clinical trials fail?. *R&D Management*, 44(2), 189-202. doi:10.1111/radm.12053

Compston, H. (2006). The expansion of mass media. In *King trends and the future of public policy*. Palgrave Macmillan.

Collins, F. S. (Summer, 2011). The Importance of Clinical Trials. *NIH MedlinePlus*. Retrieved from https://www.nih.gov/sites/default/files/about-nih/nih-director/articles/collins/importance-of-clinical-trials.pdf

Costa, L. J., Hari, P. N., & Kumar, S. K. (2016). Differences between unselected patients and participants in multiple myeloma clinical trials in US: a threat to external validity. *Leukemia and Lymphoma,* 57(12), 2827-2832, doi: 10.3109/10428194.2016.1170828

Cottingham, M. D. & Fisher, J. A. (2017). From fantasy to reality: Managing biomedical risk emotions in and through fictional media. *Health, Risk, & Society*. 19(5-6), 284-300. doi: 10.1080/13698575.2017.1350638

Dude, A., Brossard, D., Shanahan, J., Scheufele, D. A., Morgan, M., & Signorielli, N. (2011). Science on television in the 21st century: Recent trends in portrayals and their contributions to public attitudes toward science. *Communication research*, 38(6), 754-777. doi: 10.1177/0093650210384988

Egbert, L. D., Battit, G. E., Welch, C. E., & Bartlett, M. K. (1964) Reuction of postoperative pain by encouragement and instruction of patients – a study of doctor-patient rapport. T*he New England Journal of Medicine*, 270, 825-827. doi: 10.1056/NEJM196404162701606

Eggly, S., Albrecht, T. L., Harper, F. W. K., Foster, T., Franks, M. M., & Ruckdeschel, J. C. (2008). Oncologists' recommendations of clinical trial participation to patients. *Patient Education and Counseling*, 70(1), 143-148. doi: 10.1016/j.pec.2007.09.019

Elting, L. S., Cooksley, C., Bekele, B. N., Frumovitz, M., Avritschers, E. B. C., Sun, C., & D. C. Bodurka (2006). Generalizability of cancer clinical trial results: prognostic differences between participants and nonparticipants. *Cancer* 106(11). doi: 10.1002/cncr.21907

Fisher & Cottingham (2017). This isn't going to end well: Fictional representations of medical research in television and film. *Public Understanding of Science*. 26(5), 564-578. doi: .o0r.g1/107.171/0779/6039 636626521561664411339

Fogel, D. B. (2018). Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review. *Contemporary Clinical Trials Communication*, 11, 156–164. doi:10.1016/j.conctc.2018.08.001

Ha, J. F., Anat, D. S., & Longnecker, N. (2010). Doctor-patient communication: a review. T*he Ochsner Journal*, 10(1), 38-43

Hagiwara, N., Penner, L. A., Gonzalez, R., Eggly, S., Dovidio, J. F., Gaertner, S. L., West, T., Albrecht, T. L. (2013). Racial attitude, physician-patient talk time ratio, and adherence in racially discordant medical interactions. *Social Science & Medicine*, 87, 123-131. doi: 10.1016/j.socscimed.2013.03.016

Haldane, J. B. S. (1940). The mean and variance of chi2 when used as a test of homogeneity when expectations are small. *Biometrika*, 31(3/4), 346-355. doi: 10.2307/2332614

Hashim, M. J. (2017). Patient-centered communication: basic skills. *American Family Physician,* 95(1), 29-35.

Haskard, K. B., DiMatteo, M., R., & Heritage, J. (2009). Affective and instrumental communication in primary care interactions: Predicting the satisfaction of nursing staff and patients. *Health Communication,* 24(1), 21-32. doi: 0.1080/10410230802606968

Hether, H. J., Huang, G. C., Beck, V., Murphy S. T., & Valente, T. W. (2008). Entertainment-education in a media-saturated environment: Examining the impact of single and multiple exposures to breast cancer storylines in two popular medical dramas. *Journal of Health Communication*. 13(8), 808-823.

doi: 10.1080/10810730802487471.

Hwang, T. J., Carpenter, D., Lauffenburger, J. C., Wang, B., Franklin, J. M., Kesselheim, A. S. (2016). Failure of investigational drugs in late-stage clinical development and publication of trial results. *JAMA Internal Medicine*.176(12), 1826–1833. doi:10.1001/jamainternmed.2016.6008

Huntley, C. D., Salmon, P., Fisher, P. L., Fletcher, I., & Young, B. (2012). LUCAS: A theoretically informed instrument to assess clinical communication in objective structured clinical examination. *Medical Education*, 46, 267-276. doi: doi:10.1111/j.1365-2923.2011.04162.x

Institute of Medicine (2001). *Crossing the quality chasm: A new health system for the 21st century.* National Academy Press.

Jacobs, S. R., Weiner, B. J., Reeve, B. B., Weinberger, M., Minasian, L. M., & Good, M. J. (2014). Organizational and physician factors associated with patient enrollment in cancer clinical trials. *Clinical Trials*, 11(5), 565-575. doi: 10.1177/1740774514536000

Kaur, R., Sidhu, P., & Singh, S. (2016). What failed BIA 10-2474 phase I clinical trial? global speculations and recommendations for future phase I trials. *Journal of Pharmacology and Pharmacotherapeutics*, 7(3), 120-126. doi: 10.4103/0976-500X.189661

Lang, E. V. (2012). A better patient experience through better communication. J*ournal of Radiology Nursing*, 31(4), 114-119. doi: 10.1016/j.jradnu.2012.08.001

Levit, L., Balogh, E., Nass, S., & Ganz, P. A. (Eds.) (2013). *Delivering high-quality cancer care: Charting a new course for a system in crisis.* Committee on Improving the Quality of Cancer Care: Addressing the Challenges of an Aging Population; Board on Health Care Services; Institute of Medicine. National Academies Press.

Majumdar, S. R., Roe, M. T., Peterson, E. D., Chen, A. Y., Gibler, B., and Armstrong, P. W. (2008). Better outcomes for patients treated at hospitals that participate in clinical trials. *JAMA Internal Medicine*, 168(6), 657-662. doi: 10.1001/archinternmed.2007.124

McKee, S. (Writer), & Tinker, M (Director). (2008, May 1). Piece of My Heart (Season 4, Episode 13) [TV series episode]. In B. Beers, R. Corn, M. Gordon, S. Rhimes, K. Vernoff, M. Wilding (Executive Producers), *Grey's Anatomy*. Shondaland; The Mark Gordon Company; ABS Studios

Meropol, N. J., Buzaglo, J. S., Millard, J., Damjanov, N., Miller, S. M., Ridgway, C., Ross, E. A., Sprandio, J. D., & Watts, P. (2007). Barriers to clinical trial participation as perceived by oncologists and patients. *Journal of the National Comprehensive Cancer Network*, 5(8), 753-762. doi: 10.6004/jnccn.2007.0067

National Institute on Aging (NIA) (2017, Sept). What Are Clinical Trials and Studies?. Retrieved from https://www.nia.nih.gov/health/what-are-clinical-trials-and-studies

Nguyen, T. K., Nguyen, E. K., Warner, A., Louie, A. V., & Palma, D. A. (2018). Failed randomized clinical trials in radiation oncology: what can we learn?. *International Journal of Radiation Oncology Biology Physics*, 101(5), 1018-1024. doi: 10.1016/j.ijrobp.2018.04.030

Nørgarrd, B., Ammentorp, J., Kyvik, K., & Kofoed, P. (2012). Communication skills training increases self-efficacy of health care professional. *Journal of Continuing Education in the Health Professions*, 32(2), 90-97. doi: 0.1002/chp.21131

Novitzke, J. M. (2008). The significance of clinical trials. *Journal of Vascular and Interventional Neurology*, 1(1), 31

Nowalk, P. (Writer), & Ornelas, E. (Director). (2011, March 24). This is how we do it (Season 7, Episode 13) [TV series episode]. In B. Beers, R. Corn, M. Gordon, T. Phelan, J. Rater, S. Rhimes, K. Vernoff, & Mark Wilding (Executive Producers), *Grey's Anatomy*. Shondaland; The Mark Gordon Company; ABC Studios

Osborn, C. Y., Paasche-Orlow, M. K., Davis, T. C., & Wolf, M. S. (2007). Health literacy: an overlook fac-

tor in understanding HIV health disparities. *American Journal of Preventative Medicine*, 33(5), 374-378. doi: 10.1016/j.amepre.2007.07.022

Osborn, C. Y., Cavanaugh, K., Wallston, K. A., Kripalani, S., White, R. O., Elasy, T. A., & Rothman, R. L. (2013). Health literacy explains racial disparities in diabetes medical adherence. *Journal of Health Communication*. 16(3), 268-278. doi: 10.1080/10810730.2011.604388

Peck, B. M. & Conner, S. (2011). Talking with me or talking at me? The impact of status characteristics on doctor-patient interaction. *Sociological Perspectives*. 54(4), 547-567. doi: 10.1525/sop.2011.54.4.547

Quick, B. L. (2009). Coverage of the organ donation process on Grey's Anatomy: The story of Denny Duquette. *Clinical Transplantation*, 23, 788-793. doi: 10.1111/j.1399-0012.2008.00937.x

Rideout, V. (2008). Television as a health educator: A case study of *Grey's Anatomy*. Kaiser Family Foundation Report. Retrieved from https://www.kff.org/other/television-as-a-health-educator-a-case/

Rhimes, S. (Writer), & Corn, R. (Director). (2008, May 22). Freedom: Part 1 (Season 4, Episode 16) [TV series episode]. In B. Beers, R. Corn, M. Gordon, & S. Rhimes (Executive Producers), *Grey's Anatomy*. Shondaland; The Mark Gordon Company; ABC Studios.

Rodriguez, H. P., Anastario, M. P., Frankel, R. M., Odigie, E. G., Rogers, W. H., von Glahn, T., & Safran, D. G. (2008). Can teaching agenda-setting skills to physicians improve clinical interaction quality? A controlled intervention. B*ioMed Central Medical Education*, 8(3). doi: doi:10.1186/1472-6920-8-3

Shahar, T., Nossek, E., Steinberg, D. M., Rozovski, U., Blementhal, D. T., Bokstein, F., Sitt, R., Freedman, S., Corn, B. W., Kanner, A. A., & Ram, Z. (2012). The impact of enrollment in clinical trials on survival of patients with glioblastoma. *Journal of Clinical Neuroscience*, 19(11), 1530-1534. doi: 10.1016/j.jocn.2010.04.005

Sherwood, A., Brinkmann, J., & Fatone, S. (2018). Review of benefits to practitioners of using good patient-practitioner communication. J*ournal of Prosthetics and Orthotics,* 30(1), 5-12. doi: 10.1097/JPO.0000000000000165

Simpson, M., Buckman, R., Stewart, M., Maguire, P., Lipkin, M., Novack, D., & Till, J. (1991). Doctor-patient communication: The Toronto consensus statement. *British Medical Journal*, 303(6814), 1385-1385. doi: 10.1136/bmj.303.6814.1385

Song, L., Ji, Y., & Nielsen, M. E. (2014). Quality of life and health status among prostate cancer survivors and noncancer population controls. *Urology*, 83(3), 658-663. doi: 10.1016/j.urology.2013.12.009

Soones, T. N., Lin, J. L., Wolf, M. S., O'Conor, R., Martynenko, M., Wisnivesky, J. P., & Federman, A. D. (2017). Pathways linking health literacy, health beliefs, and cognition to medication adherence in older adults with asthma. *Journal of Allergy and Clinical Immunology*. 139(3), 804-809. doi: 10.1016/j.jaci.2016.05.043

Stensland, K. D., McBride, R. B., Latif, A., Wisnivesky, J., Hendricks, R., Roper, N., Boffetta, P., Hall, S. J., Oh, W. K., & Galsky, M. D. (2014). Adult cancer clinical trials that fail to complete: An epidemic?, *JNCI: Journal of the National Cancer Institute*, (106)9, doi: 10.1093/jnci/dju229

Stewart, M., Brown, J. B., Donner, A., McWhinney, I. R., Oates, J., Weston, W. W., & Jordan, J. (2000). The impact of patient-centered care on outcomes. *The Journal of Family Practice*, 49(9), 796-804.

Stone, N. B., Kerr, C. E., Jacobson, E., Conboy, L. A., & Kaptchuk, T. J. (2005). Patient expectations in placebo-controlled randomized clinical trials. *Journal of Evaluation in Clinical Practice. 11*(1), 77-84. doi: 10.1111/j.1365-2753.2004.00512.x

Street, R. L., Liu, L., Farber, N. J., Chen, Y., Calvitti, A., Zuest, D., Gabuzda, M. T., Bell, K., Gray, B., Rick, S., Ashfaq, S., & Agha, Z. (2014). Provider interaction with the electronic health record: The effects on patient-centered communication in medical encounters. *Patient Education and Counseling*, 96(3), 315-319. doi: 10.1016/j.pec.2014.05.004

Stuart, E. A., Bradshaw, C. P, & Leaf, P. J. (2015). Assessing the generalizability of randomized trial results to target populations. *Prevention Science*, 16(3), 475-485, doi: 10.1007/s11121-014-0513-z

Stunkel, L. and Grady, C. (2010). More than the money: A review of the literature examining healthy volunteer motivations. *Contemporary Clinical Trials*, 32(3), 342-352. doi: 10.1016/j.cct.2010.12.003

Teutsch, C. (2003). Patient-doctor communication. *Medical Clinics of North America*, 87(5). 1115-1145. doi: 0.1016/S0025-7125(03)00066-X

Tsoh, J. Y., Sentell, T., Gildengorin, G., Le, G. M., Chan, E., Fung, L. C., Pasick, R. J., Stewart, S., Wong, C., Woo, K., Burke, A., Wang, J., McPhee, S. J., & Nguyen, T. T. (2016). Healthcare communication barriers and self-rated health in older Chinese American immigrants. *Journal of Community Health*, 41(4), 741-752. doi: 10.1007/s10900-015-0148-4.

U.S. Department of Health and Human Services (2018a). Table A-18a. Age-adjusted percent distribution (with standard errors) of length of time since last contact with doctor or other health care professional among adults aged 18 and over, by selected characteristics: United States, 2018. *Summary Health Statistics: National Health Interview Survey,* 2018. Retrieved from https://ftp.cdc.gov/pub/Health_Statistics/NCHS/NHIS/SHS/2018_SHS_Table_A-18.pdf

U.S. Department of Health and Human Services (2018b). Table C-8a. Age-adjusted percent distribution (with standard errors) of length of time since last contact with a doctor or other health care professional for children under age 18 years, by selected characteristics: United States, 2018. *Summary Health Statistics: National Health Interview Survey*, 2018. Retrieved from https://ftp.cdc.gov/pub/Health_Statistics/NCHS/NHIS/SHS/2018_SHS_Table_C-8.pdf

U.S. Food and Drug Administration (2017). 2015-2016 Global Participation in Clinical Trials Report. Retrieved from https://www.fda.gov/media/106725/download

Unger, J. M., Cook, E., Tai, E., & Bleyer, A. (2016). The Role of Clinical Trial Participation in Cancer Research: Barriers, Evidence, and Strategies. *American Society of Clinical Oncology educational book. American Society of Clinical Oncology*. Annual Meeting, 35, 185–198. doi:10.14694/EDBK_156686

Wiering, B., de Boer, D., Krol, M., Wieberneit-Tolman, H., & Delnoij, D. (2018). Entertaining accurate treatment expectations while suffering from chronic pain: An exploration of treatment expectations and the relationship with patient-provider communication. *BMC Health Services Research*, 18(1). doi: 10.1186/S12913-018-3497-8

Wilding, M. (Writer), & McKidd, K. (Director). (2011, February 3). Don't Deceive Me (Please Don't Go) (Season 7, Episode 13) [TV series episode]. In B. Beers, R. Corn, M. Gordon, T. Phelan, J. Rater, S. Rhimes, K. Vernoff, & M. Wilding (Executive Producers), *Grey's Anatomy*. Shondaland; The Mark Gordon Company; ABC Studios

Wilson, E., Chen, A. H., Grumbach, K., Wang, F., Fernandez, A. (2005). Effects of limited English proficiency and physician language on health care comprehension. *Journal of General Internal Medicine*. 20(9), 800-806. doi: 10.1111/j.1525-1497.2005.0174.x

Winkfield, K. M., Phillips, J. K., Joffe, S., Halpern, M. T., Wollins, D. S., & Moy, B. (2018). Addressing financial barriers to patient participation in clinical trials: ASCO policy statement. *Journal of Clinical Oncology*, 36(33), 3321-3329. doi: DOI: 10.1200/JCO.18.01132

Wouters, O. J., McKee, M., & Luyten, J. (2020). Estimated research and development investment needed to bring a new medicine to market, 2009-2018. *JAMA*. 323(9), 844–853. doi:10.1001/jama.2020.1166

Zolnierek, K. B. H. & DiMatteo, M. R. (2009). Physician communication and patient adherence to treatment: a meta-analysis. *Medical Care*, 47(8), 826-834. doi: 10.1097/MLR.0b013e31819a5acc

Appendix A: Screened television episodes by year

Title	Season	Episode	Year
Malcolm in the middle	6	20	2005
Grey's Anatomy	4	13	2008
Grey's Anatomy	4	14	2008
Grey's Anatomy	4	15	2008
Grey's Anatomy	4	16	2008
Grey's Anatomy	4	17	2008
Law and Order: SVU	10	1	2008
House M.D.	5	3	2008
House M.D.	5	11	2008
House M.D.	5	14	2008
Grey's Anatomy	5	13	2009
All Saints	12	6	2009
American Dad!	4	8	2009
Grey's Anatomy	7	13	2011
Grey's Anatomy	7	16	2011
Grey's Anatomy	7	17	2011
Grey's Anatomy	7	19	2011
Grey's Anatomy	7	22	2011
Grey's Anatomy	8	7	2011
Lost Girl	2	15	2012
Archer	3	5	2012
It's Always Sunny in Philadelphia	9	8	2013
Drop Dead Divas	5	1	2013
Murdoch Mysteries	7	5	2013
Maniac	1	1	2018
Maniac	1	2	2018
Maniac	1	3	2018
Maniac	1	4	2018
Maniac	1	5	2018
Maniac	1	6	2018
Maniac	1	7	2018
Maniac	1	8	2018
Maniac	1	9	2018
Maniac	1	10	2018
Grey's Anatomy	15	22	2018
Grey's Anatomy	15	22	2018
Grey's Anatomy	15	22	2018
Grey's Anatomy	15	22	2018

Appendix B: Huntley et al. (2012) LUCAS Scoring Descriptors

LIVERPOOL UNDERGRADUATE COMMUNICATION ASSESSMENT SCALE (LUCAS): SCORING DESCRIPTORS

Item	Competent	Unacceptable
A) Greeting & introduction	i) greets patient, ii) provides full name, iii) job title, iv) brief explanation why s/he is approaching the pt. *E.g." Good morning, my name's Jon Dough and I'm a first year medical student from Liverpool University. The doctor has asked me to come and speak with you today."*	Omission of **any** of elements i-iv
B) Identity check	i) check of pt's full name (if consultation is with a relative or carer, the candidate should check the name of the person with whom they are speaking and their relationship to the pt), ii) check of one other identifier (e.g. pt's D.O.B., address etc.)	Omission of **either** i) **or** ii)

	Competent	Borderline	Unacceptable
C) Audibility and clarity of speech	Speech is clear or mostly clear; moderates voice or uses repetition when necessary and the patient is likely to hear all key points.	Speech is somewhat unclear in places; attempts to moderate voice or use repetition but the patient may not hear some key points.	Speech is mostly unclear i.e. voice too quiet or enunciation too fast or indistinct for patient to hear; fails to moderate voice or use repetition for patient to hear.
D) Non-verbal behaviour (NVB)	NVB facilitates (i.e. calm and confident) or largely facilitates engagement (i.e. may appear a little uneasy but this unlikely to impact adversely on the patient's engagement with the consultation).	NVB is awkward and likely to be distracting to the patient and limit his/her engagement with the consultation at times.	NVB is inappropriate to healthcare setting, or likely to be disconcerting for the patient to the extent that it prevents or substantially disrupts the patient's engagement with the consultation.
E) Questions, prompts and/or explanations (QPE) (N.B. this item is not to assess the medical content of history taking, which is rated in other OSCE stations, or on separate mark sheets)	QPEs address patient's key needs, feelings and concerns and any omissions are minor; QPEs are mostly easy to understand (e.g. jargon is used sparingly and explained).	Attempts to address the patient's needs, views and feelings, but QPEs are inadequate or incomplete so *key* needs are only *partially* covered; QPEs somewhat difficult to understand.	Fails to explore or explain issues crucial in addressing the patient's needs, views and feelings; QPEs very difficult to understand (e.g. poor choice of words and little or no attempt to rephrase) to the extent that the patient's key needs are not addressed.
F) Empathy and responsiveness	Responsive and sensitive to the patient's needs, views and feelings though there may be room for improvement (e.g. responses might be slightly perfunctory at times); makes good or adequate use of reflection/verbal acknowledgement; evidence of care and concern for the patient as a person.	Attempts to acknowledge or be sensitive to patient's main needs, views and feelings but responses are generally incomplete or inadequate (e.g. responds in a way that is obviously cursory or superficial); approach to patient appears distant or distracted.	Little or no acknowledgement of, or sensitivity to, the patient's needs, views and feelings (e.g. ignores patient's main concerns, or responds in a way that is uncaring when patient voices his/her views and feelings); approach to patient appears cold or indifferent.

Appendix C: Adapted Huntley et al. (2012) LUCAS Scoring Descriptors

	Score		
	0	1	2
A) Non-verbal behavior (NVB)	NVB is inappropriate to healthcare setting, or likely to be disconcerting for the patient to the extent that it prevents or substantially disrupts the patient's engagement with the consultation.	NVB is awkward and likely to be distracting to the patient and limit his/her engagement with the consultation at times.	NVB facilitates (i.e. calm and confident) or largely facilitates engagement (i.e. may appear a little uneasy but this unlikely to impact adversely on the patient's engagement with the consultation).
B) Questions, prompts, and explanations (QPEs)	Fails to explore or explain issues crucial in addressing the patient's needs, views and feelings; QPEs very difficult to understand (e.g. poor choice of words and little or no attempt to rephrase) to the extent that the patient's key needs are not addressed.	Attempts to address the patient's needs, views and feelings, but QPEs are inadequate or incomplete so key needs are only partially covered; QPEs somewhat difficult to understand.	QPEs address patient's key needs, feelings and concerns and any omissions are minor; QPEs are mostly easy to understand (e.g. jargon is used sparingly and explained).
C) Empathy and responsiveness	Little or no acknowledgment of, or sensitivity to, the patient's needs, views and feelings (e.g. ignores patient's main concerns, or responds in a way that is uncaring when patient voices his/her views and feelings); approach to patient appears cold or indifferent.	Attempts to acknowledge or be sensitive to patient's main needs, views and feelings but responses are generally incomplete or inadequate (e.g. responds in a way that is obviously cursory or superficial); approach to patient appears distant or distracted.	Responsive and sensitive to the patient's needs, views and feelings though there may be room for improvement (e.g. responses might be slightly perfunctory at times); makes good or adequate use of reflection/verbal acknowledgment; evidence of care and concern for the patient as a person.

Identification of genes associated with Autism Spectrum Disorder that contribute to social behavior defects in *Caenorhabditis elegans*

Kaelin Rubenzer
The University of Texas at Austin
Texas Undergraduate Research Journal

ABSTRACT

Autism spectrum disorder (ASD) is a neurodevelopmental disorder often characterized by significant problems with social skills. Human genetic studies have associated ASD with mutations in over 200 genes; however, it is unknown which ASD candidate genes are singly causal versus coincidentally associated with social deficits. Testing for causality of ASD-related phenotypes in such a large list of genes is challenging. We sought to leverage the convenient behavioral genetics of *Caenorhabditis elegans* to test for social deficits in orthologs of ASD-associated genes. The common wild-type lab strain display only weak social behavior due to a background mutation in *npr-1*, but many natural wild-isolate strains are strongly social. Using wild isolate strains that contained unique, naturally-occurring variants predicted to alter function of conserved ASD candidate genes, we evaluated levels of social feeding behaviors for defects in the form of bordering and clumping. We discovered that strains with predicted change-of-function mutations in the orthologs of human genes DYRK1A, SHANK3, NLGN1 and SETD5 demonstrated significantly lower levels of social behavior compared to other social strains with normal sequence of these genes. Analysis of the evolutionary conservation of ASD implicated genes was also studied and characterized to better understand critical regions that could be especially sensitive to naturally occurring variations. Identification of genes and their critical regions that may predict ASD phenotypes could help medical professionals and genetic counselors to identify at-risk families, while also providing useful insight to explain causality in the biological foundations of ASD.

INTRODUCTION

According to the Autism and Developmental Disabilities Monitoring (ADDM) Network, approximately 1 in 59 children are diagnosed with autism spectrum disorder (ASD). ASD is a pervasive developmental disorder characterized by significant problems with social, behavioral, and communication skills. People diagnosed with ASD often will have issues with social reciprocity, will demonstrate restricted and repetitive patterns of behaviors or interests, and may have unusual sensory sensitivity. Speech deficits, intellectual delays, and other conditions such as sleep problems, epilepsy and mental illnesses are also common in ASD patients. The variety and severity of symptoms in ASD differ greatly in each person, making it an extremely complex disorder to characterize and study. The wide range of ASD symptoms highly suggests that ASD is caused by several contributing factors, likely a complex interaction between environmental and genetic influences. Non-genetic factors, such as parental age and birth complications, likely synergize with genetic factors to create the numerous complex phenotypes of ASD.

The heterogeneous expression of ASD phenotypes has prompted rich research emerging from multiple distinct disciplines and areas of intervention and care. ASD was first formally recognized and de-

scribed by Dr. Leo Kanner from John Hopkins Hospital in 1942, who described a 5-year-old boy who preferred social isolation, demonstrated significant speech delay, fixated on repetitive tasks and toys, and was unusually inflexible about routine changes. These now-easily recognizable ubiquitous red-flags for ASD became the earliest documented clinical definition and criteria for ASD diagnosis, which paved the way for more accurate recognition and characterization of this group of disorders as time progressed. Early ASD research focused on observing ASD in all age groups, and were able to characterize language and expression challenges, develop behavioral therapy strategies and special education programs, and address the additional stressors to the family dynamic. Although contributions from social workers, special education teachers, behavioral psychologists and other specialists have contributed greatly to our multidimensional understanding of ASD and its symptom management, the need for research into the neurobiological foundations of ASD remain imperative for developing more sophisticated improvements in diagnosis and treatment strategies.

Genes Implicated in ASD

Currently over 1,000 genes, regarded as ASD-candidate genes, have been associated with this disorder by heritable genetic variation and de novo mutations. The leading hypothesis for the neurophysiological foundations of ASD suggests that defects in thalamocortical circuitry cause an imbalance of excitatory and inhibitory neuronal activity, leading to the altered neurodevelopment associated with ASD and other similar intellectual disorders. In a large-scale exome sequencing study by Satterstrom et al. in 2020, evaluation of 35,584 samples (11,986 with ASD) revealed 102 genes with a range of expression patterns and functions were implicated in risk for ASD, with 49 of them also being associated with other developmental delays. These genes were often expressed early in neurodevelopment and are either responsible for the regulation of other genes or for synaptic maintenance, communication, and/or architecture, indicating the range of global and neurophysiological alterations that potentially cause ASD phenotypes.

However, a majority of these associations have not been scientifically confirmed. Very little is known about which ASD candidate genes are causal versus coincidentally associated with various ASD behavioral phenotypes, due to the sheer size of the list of candidate genes. Identification of the genes that may predict the onset of ASD phenotypes could help medical professionals and genetic counselors to identify at-risk families, while also providing useful insight to explain causality in the biological foundations of ASD. To address the issue of determining which ASD candidate genes are causal for ASD phenotypes, we sought to leverage the convenient behavioral genetics and high throughput system of the *Caenorhabditis elegans* (*C. elegans*) animal model to evaluate what ASD gene variants could be responsible for common ASD phenotypes.

C. elegans as an ASD model organism

The humble, free-living roundworms known as *C. elegans* have increasingly gained traction in the research community for being a valuable animal model for multiple human disorders and genetics studies. *C. elegans* are self-fertilizing hermaphrodites and clonal, allowing for easy preservation of established gene lines. The life-span of *C. elegans* averages around 2 weeks, which eliminates the issue of slow neurodevelopment seen in rodent ASD studies. Maintenance of hundreds of worms is extremely simple relative to other animal model options, allowing for numerous genetic lines to run at once. The powerful genetics and ease of behavioral observation make *C. elegans* a good potential candidate for ASD research investigating the causal nature of multiple implicated genes.

Genes that have been studied and implicated extensively in other ASD animal models often have equivalent counterparts in *C. elegans*, with the neuronal role of several important ASD candidate genes actually first being discovered and characterized in *C. elegans*. Current biological mechanism studies of ASD have implicated the synaptic adhesion molecules Neurexin and Neuroligin, which are known to dramatically alter the development and function of synapses in the brain. Previous *C. elegans* studies have highlighted the importance of retrograde synaptic signaling of α-Neurexin for regulation of synaptic

transmission, which is a common area of interest for neurophysiological ASD researchers (Tong et al. 2017). SHANK, an intracellular scaffolding protein also heavily implicated in ASD, has three redundant SHANK genes in mammals, which has limited the success of animal model mutant knockouts due to an inconsistent compensatory response and possible overlapping functions of the SHANK protein variants. *C. elegans* only have a single equivalent SHANK gene, which allow researchers to study the SHANK protein's effects on synaptic transmission in a more simple, controlled model.

Although the usefulness of *C. elegans* as an animal model has been proven in small scale ASD relevant molecular studies and other human disease models, the possibility of developing an ASD model is only now gaining attention. Unpublished research from the Rankin lab in BC Vancouver has begun exploring the potential for *C. elegans* to be used as an ASD genetic model, but has focused solely on morphology, sensory sensitivity, habituation responses, and locomotive behavior. Another study has identified highly conserved human ASD-associated missense variants and utilized a CRISPR/Cas9-mediated homology-directed knock-in strategy in N2 background worms to generate missense mutants and found promising defects in morphology, locomotion, and fecundity (Wong et al 2019). Although these structural and behavioral changes are relevant to some characteristic features of ASD, no current research is being done to assess potential changes in social behavior, one of the most characteristic red-flags for an ASD diagnosis.

For the last 5 decades of research using domesticated *C. elegans* strains, the laboratory strain N2 were described as having solitary feeding behavior when food was plentiful, and only aggregated temporarily when food was scarce. Further research involving new wild-isolate strains from different geographical regions revealed significant variations in each strain's social feeding behavior, which prompted further forward genetics studies. It was discovered that domestication of the lab wild-type N2 strain from Bristol, UK, had inadvertently been selected for solitary feeding behavior, due to a spontaneous genetic polymorphism gain of function (gf) mutation at a single locus, called *npr-1*. Because many domesticated lab strains contained this background spontaneous mutation, the potential to study social behavior in *C. elegans* was generally overlooked for many years.

Npr-1 is now known to encode a G-Protein-coupled neuropeptide Y receptor homolog, and has been implicated in the multiple behavioral responses of *C. elegans* to bacterial food, namely aerotaxis and aggregation. Inducing a loss-of-function mutation in the *npr-1* gene was able to restore social feeding behavior of N2 strains back to averages observed in wild-isolate worms. Wild-isolate worms without background *npr-1* (gf) mutations demonstrate predictable social feeding patterns, in the form of bordering and clumping. Because defects in social behavior are a ubiquitous trait in ASD, we sought to leverage the advantages of the *C. elegans* model to investigate which ASD implicated genes are causal for social defects.

Evolutionary Conservation of ASD Candidate Genes

Comparing ASD candidate gene sequences and their respective ortholog counterparts in other invertebrate and mammal species allows us to identify important, highly-conserved segments in the genes of interest. Determining these critical regions will allow researchers to filter out the "junk" or unimportant intron segments of a particular gene, so the development of mutant animal models can solely focus on impactful mutations in the exon portion of an ASD candidate gene.

A helpful analogy for highlighting the importance of identifying critical regions versus intron segments is likening a gene sequence to a family recipe. Following any standard baking recipe, there are some ingredients that are interchangeable and there are some that are necessary for the identity of the dish to remain true. If you're baking a loaf of bread, you'll need flour, some proportion of yeast, salt, and filtered water. Any variation to these ingredients or proportions will often lead to a disappointing final product. However, many recipes allow for some additional add-ins, or components that can be substituted for similar ingredients.

Similarly, the deliberate organization and sequence of different nucleotides provides the cell an informa-

tional "recipe" for building a final, functional gene product, most often a protein. Each triplet sequence of nucleotides encodes a specific amino acid, which carry distinct chemical properties and charges due to their variable side chain. Strings of these amino acids and their charged side chains fold, bond, and form complex three-dimensional configurations to compose functional proteins. Proper protein folding requires that the right amino acids are in correct spatial orientation, otherwise, protein structure and function can be dramatically altered. These are like our flour and essential components: the gene needs to highly conserve these regions to preserve the integrity of the final product. However, there are segments of this peptide sequence that are relatively unimportant for intramolecular bonding, and therefore may be interchangeable without affecting the overall structure and function of the final desired gene product. Similar to the more flexible parts of a recipe, these small alterations (such as an amino acid swap to another with a similar charge) will not dramatically alter the final gene product, so researchers can choose to focus on more significant possible variants.

Continuing with the family recipe analogy, evolutionary conservation of a particular gene can be likened to passing on the recipe to a neighbor. As time progresses and the informational code of a recipe is passed into different hands, there are portions of the recipe that must stay consistent and other components that are open to embellishment. Similarly, the conservation of an ASD gene in different animal genes over the evolutionary timespan will likely have significant variation, but there will likely be strictly conserved regions that are imperative for the basic gene function. We hope to identify these highly conserved regions by comparing multiple gene sequences for evolutionarily conserved ASD genes from different species.

METHODOLOGY

C. elegans maintenance

Worms are grown and maintained on Nematode Growth Medium (NGM)-agar filled, 60-mm polystyrene Petri dishes seeded with the food source, Escherichia coli strain OP50. Separate genetic strains are maintained by moving 5-10 L4 or young adult hermaphrodites to a freshly seeded plate every 3-4 days and then kept at room temperature in sealed bins until experimentation. The L4 or young adult worms used for propagation are chosen at random and must be independent from clumps to mitigate potential genetic drift due to an inadvertent bottleneck effect.

Selection of C. elegans ASD strains

Strains utilized in this study will each have a single, known natural high-impact variant in a single conserved ASD candidate Category 1 gene. The summarized search process is detailed in Figure 1.

Identifying highly-conserved regions of evolutionary preserved human ASD gene-orthologs

After compiling a starting list for category 1 ASD implicated genes, the molecular evolution of their protein-coding regions was analyzed across invertebrate and vertebrate species. Orthologous sequences of each ASD gene was obtained via the Worm Base Homology search engine and the protein sequence database BLAST. Evolutionarily conserved peptide sequences in the form of FASTA sequences from different animal species was then entered and analyzed through the multiple sequence alignment program Clustal Omega. Proper alignment of multiple orthologous peptide sequences allowed proper identification of highly conserved regions in each gene. Human ASD implicated genes and their orthologous counterparts were evaluated for average amino acid length and for evidence of evolutionary conservation. For a region to be considered highly conserved, we determined that spatially aligned sequences must contain the identical sequences for at least five consecutive amino acids. Regions considered moderately conserved were sequences that contained similar sequences with occasional substitutions for other amino acids that had similar chemical properties in all species. Regions that showed significant conservation but differed significantly in one or more groups were denoted for further investigation.

Social behavior assays

This project will focus on using various natural isolate worm strains from various geographical locations, which all, in theory, demonstrate natural social feeding behavior in the form of bordering and clumping. Each ASD worm strain will be evaluated for defects in social feeding behavior, which will ideally indicate whether that specific variant is causal for social defects or not.

Wild isolate *C. elegans* demonstrate social bordering behavior by aggregating on the edge of the bacterial lawn of their culture plates. They also demonstrate social clumping behavior by aggregating in large piles of 3 or more individuals. Fifty L4-stage worms will be transferred to fresh seeded agar plates. After 24 hours, the percentage of worms on the border and number of clumps will be calculated. Eight replicates will be performed for each isolate. Average percentage values will be compared with ANOVA with Tukey's posthoc comparison.

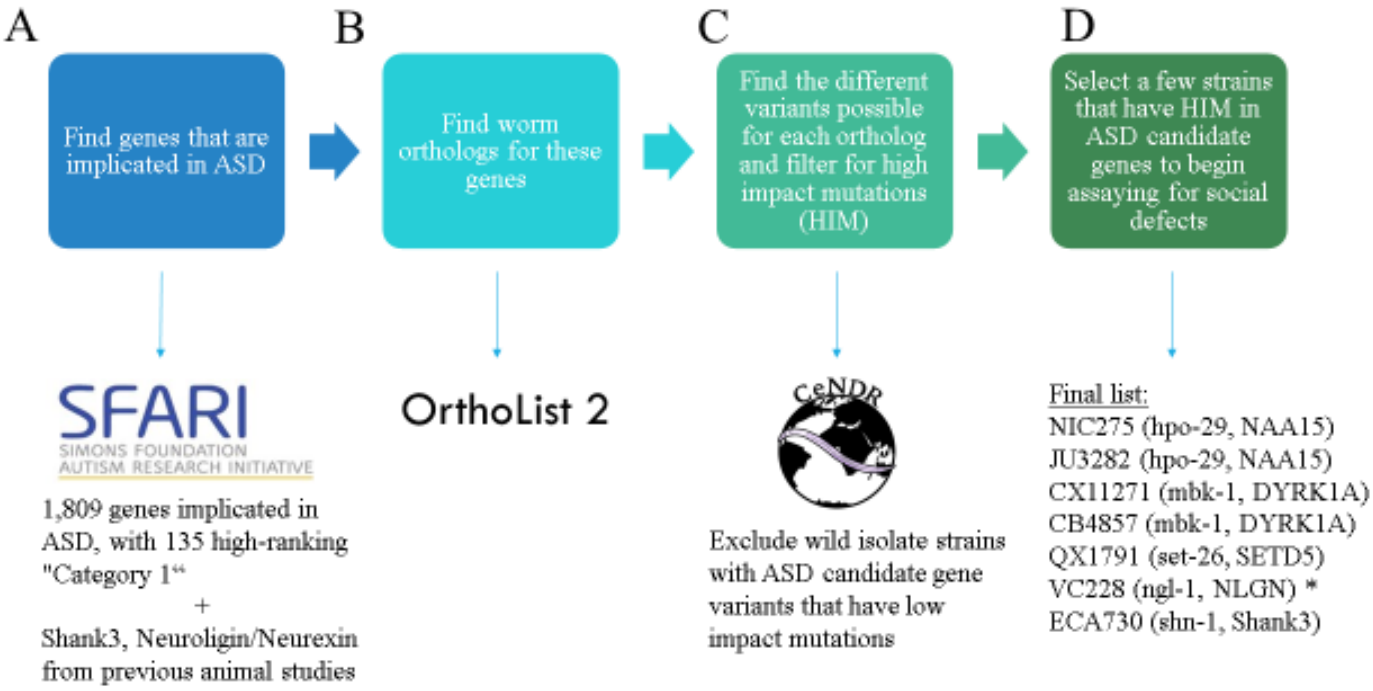
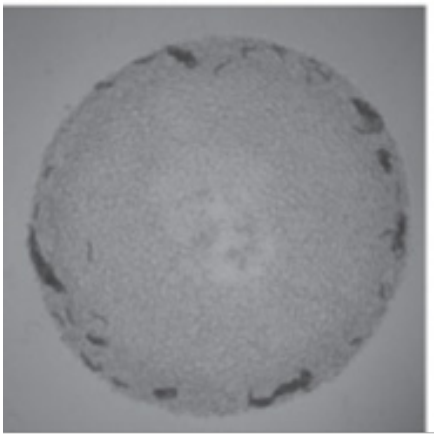
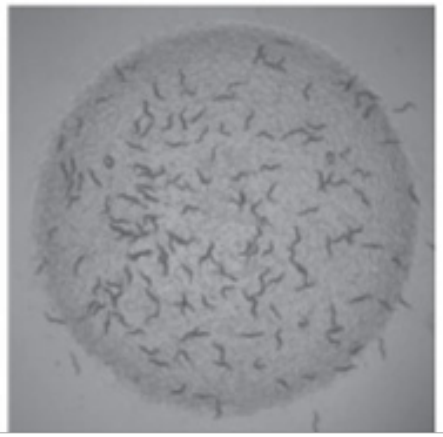


Figure 1: Process for selecting ASD worm strains. (A) ASD candidate genes evaluated in this study have been implicated through other ASD animal studies or through SFARI, an evolving database and resource for ASD researchers. SFARI has released a continuously updated list categorizing and prioritizing ASD implicated genes into tiered lists, with "Category 1 genes" as genes that have been extensively researched and consistently associated with ASD. **(B)** Each Category 1 gene was then manually evaluated for an equivalent *C. elegans* ortholog using the OrthoList2 database created by the Dan Shaye lab. Any preliminary Category 1 ASD genes that did not have an equivalent *C. elegans*-human ortholog was excluded from this study. **(C)** Once possible *C.elegans* ASD gene-orthologs were identified, these orthologs were individually processed through the *Caenorhabditis elegans* Natural Diversity Resource (CeDNR) database and variant browser to identify different variants possible for each ortholog. Because many of these possible variants in geographically distinct worm populations were observed in naturally-occurring, viable communities, many possible variants in these genes of interest were considered low-impact, meaning the genetic variation often did not dramatically alter the general fitness or survival of the worms. To ensure clear experimental results that display an obvious causal relationship between a single gene variant and social behavior defect, only high-impact mutations (HIM) were selected. **(D)** Seven unique worm strains containing high-impact natural variants in Category 1 ASD-implicated genes, most with an N2 genetic background, were selected for this study.

A





B

C

Strain Name	Human ASD Candidate Gene	C. *elegans* Ortholog	Ortholog Mutation
NIC275	NAA15	Hpo-29	Stop variant, p.Gln374
JU3283	NAA15	Hpo-29	Stop variant, p.Gln374
CB4857	DYRK1A	Mbk-1	Splice acceptor/intron variant on X chromosome at 14,820,979
CX11271	DYRK1A	Mbk-1	Splice acceptor/intron variant on X chromosome at 14,820,979
QX1791	SETD5	Set-26	Stop variant/conservative in frame insertion, p.Gln683_Gln684
VC228 (N2 background)	NGLN1	Ngl-1	Allele insertion of 334bp, deletion of 2341bp X:13626899..13629239
ECA730	SHANK3	Shn-1	Missense variant, p.Thr1066Ile
N2 (control)	-	-	-
CB4932 (positive control)	-	-	-
RC301 (positive control)	-	-	-

Figure 2: Social feeding behavior of clumping and bordering evaluated in ASD strains. (A) Certain wild isolate strains are more social (left) than others (right) defined by the degree that they cluster on the border of a bacterial lawn (left). From Macosko et. al 2009 **(B)** Social worms tend to form large, multi-worm piles , formally described as the behavior clumping. **(C)** Strains used for social behavior assays. Seven distinct worm strains carrying a naturally-occurring variant in a Category 1 ASD gene were selected for experimentation. The lab wild-type N2 strain is used as an experimental control due to its well-documented absence of social feeding behavior due to a genetic polymorphism in *npr-1*. Tauton, England strain CB4932 and Freiburg, Germany strain RC301 demonstrate unusually high social feeding behavior, and were used as positive controls. Use of VC228 strain (variant in nlg-1 ortholog with N2 genetic background) was used instead of another nlg-1 strain with a wild-isolate background was due to the convenience of this strain already being available in the Pierce lab.

RESULTS

Bordering and clumping assays reveal social behavior defects in several ASD strains

To assess if each strains particular ASD gene variant was causal for a social behavior defect, each strain underwent eight distinct trials to assess bordering and clumping behavior. Behavior assays were performed mid-day and double-blinded. Due to the time-consuming nature of planning and performing each behavior assay, social behavior data was collected from June 2019 through late November. Containers holding the experimental plates were consistently placed in an area with moderate natural light, minimal exposure to air conditioning vents, and little disruptive movement for the 24-hour waiting period. Experimental plates were scored quickly and then discarded.

Different ASD strains displayed variability in percentage of bordering and clumping (Figure 3). Interestingly, not all strains that displayed defects in one social behavior had defects in the latter.

significant. N2 wild-type worms demonstrated low bordering behavior (40.34%±18.79) and no clumping behavior, which is consistent with previous Pierce lab studies showing bordering behavior within a range of 40-60% and no clumping. Positive control strains RC301 and CB4932 demonstrated high bordering behavior (69.45%±21.62 and 76.64%±20.10 respectively) and consistently high clumping (40.08%±18.79 and 54.02%±17.91 respectively) as expected.

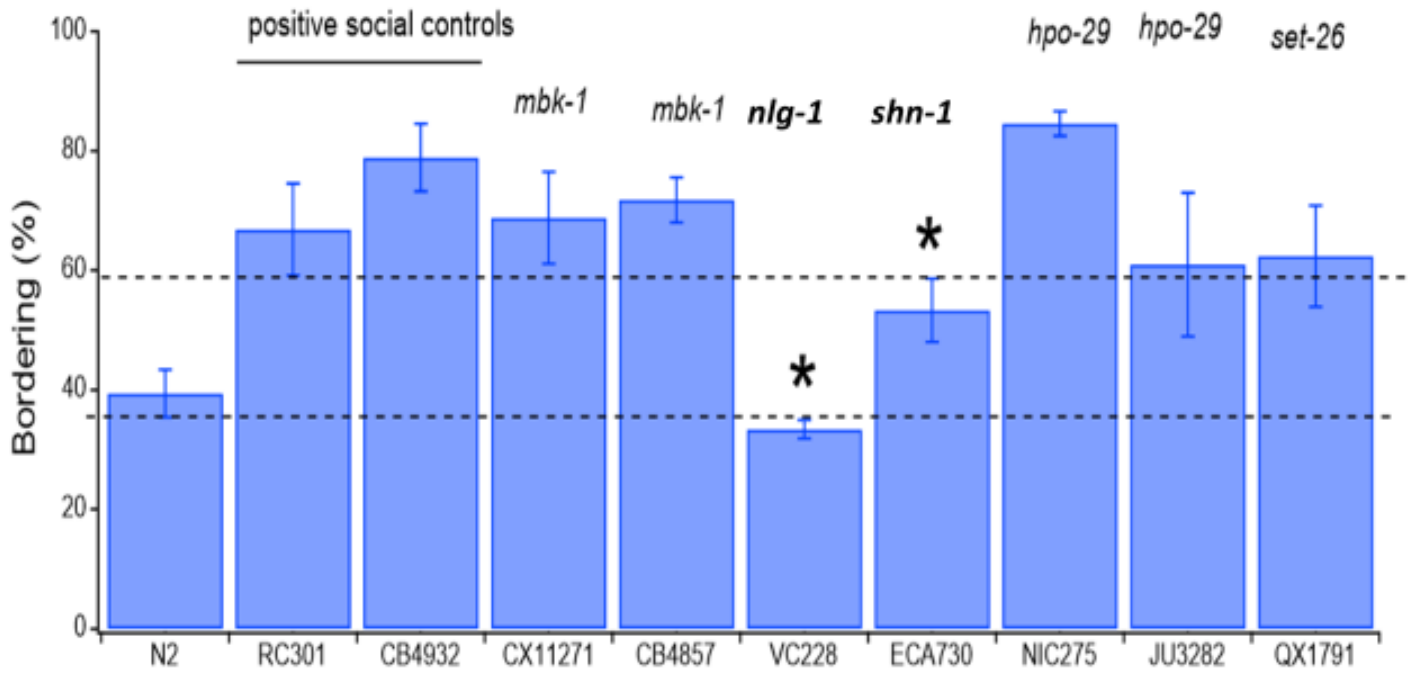
Multiple sequence analysis reveals highly conserved gene areas in NGLN1, SHANK3, NAA15 and DYRK1A, evolutionarily conserved ASD candidate genes

Sequences of human ASD implicated genes and their orthologous counterparts from diverse animal species were spatially aligned for multiple sequence analysis (MSA). Evolutionary conservation analysis of five ASD candidate genes revealed multiple highly conserved regions in some genes studied. (Figure 4). Gonnet PAM 250 distance matrices were used to score each aligned peptide sequence and to determine their similarity. Replication of the entire amino acid sequence and analysis for each gene and equivalent orthologs is provided in Supplemental Figures 1-5.

A

Strain Name, *C. elegans* Ortholog	Bordering Average (%)	Bordering Standard Deviation	Clumping Average (%)	Clumping Standard Deviation
N2 (control)	40.34	12.90	0	0
RC301 (positive social control)	69.45	21.62	40.08	18.79
CB4932 (positive social control)	76.64	20.10	54.02	17.91
CX11271, *mbk-1*	54.43	8.63	30.88	15.36
CB4857, *mbk-1*	71.68	12.33	19.47	5.10
VC228, *nlg-1*	33.94	4.03	0	0
ECA730, *shn-1*	55.62	15.13	20.33	14.49
NIC275, *hpo-29*	84.01	5.63	51.29	10.26
JU3282, *hpo-29*	60.93	24.06	39.83	36.55
QX1791, *set-26*	60.15	21.81	15.94	15.94

B



C

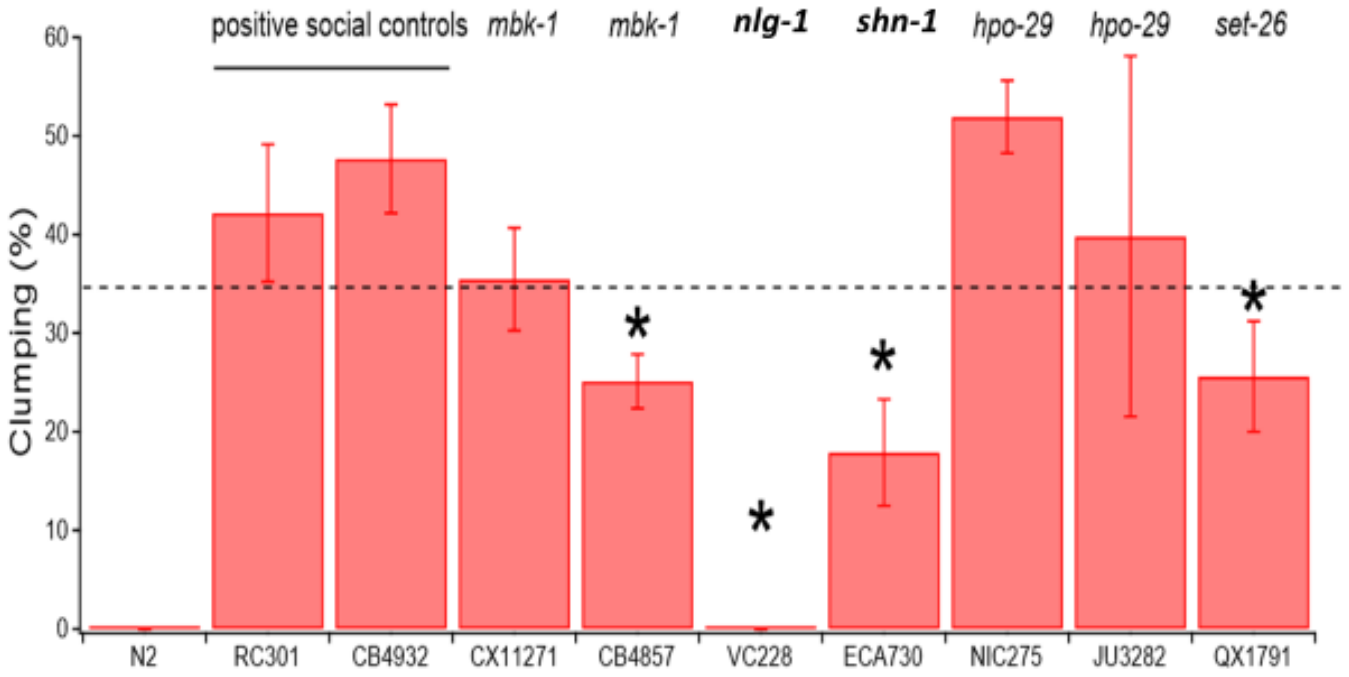


Figure 3: Strains carrying variants in the orthologs for human genes DYRK1A, SHANK3, NLGN1 and SETD5 display significant social behavior defects. (A) Table summarizing bordering and clumping data average percentages and standard deviation for each strain. **(B)** ASD strains VC228 (nlg-1) and ECA730 (*shn-1*) display significantly decreased bordering behavior compared to other wild-isolate strains and positive controls CB4932 and RC301 (33.94%±4.03 and 55.62%±15.13, respectively). Strains JU3282 (*hpo-29*) and QX1791 (set-26) also displayed a decrease in bordering behavior (60.93%±24.06 and 60.15%±21.821 respectively), but it was not deemed significant. **(C)** Strains CB4857 (*mbk-1*), VC228 (nlg-1), ECA730 (*shn-1*), and QX1791 (set-26) demonstrated decreased clumping behavior (19.47%± 5.10, 0%±0, 20.33%±14.49, and 15.94%±15.94, respectively) compared to other wild-isolate and positive controls. Strain CX11271 (*mbk-1*) also demonstrated a decrease in clumping behavior (30.88%±15.36) but it was not deemed.

A


C. elegans ----------NYLPVMVIVHGEEYGWGTGNAFNGTTLAAYGHIIVVTLNYR 182
D. melanogaster----------KQYAVLVYLHGESFEWNSGNPYDGSVLSSYGEVIVVTVNYR 315
H. sapiens TDDLGDNDGAEDEDIRDSGGPKPVMVYIHGGSYMEGTGNLYDGSVLASYGNVIVITVNYR 251
R. norvegicus ----------DIRDSGGPKPVMVYIHGGSYMEGTGNLYDGSVLASYGNVIVITVNYR 231
M. musculus TDDLGDNDGAEDEDIRDSGGPKPVMVYIHGGSYMEGTGNLYDGSVLASYGNVIVITVNYR 231
C. elegans LGVFGFLGRCESSSCSGNSGISDLVSALTMLNVILPSFGGDSKSVTLAGWGSGASLVSLL 242
D. melanogaster LGVLGFLRPSIDAHNIANYALLDQIAALHWIKENIEAFGGDNSRVTLMGHSTGAACVNYL 375
H. sapiens LGVLGFLSTGD-QAAKGNYGLLDLIQALRWTSENIGFFGGDPLRITVFGSGAGGSCVNLL 310
R. norvegicus LGVLGFLSTGD-QAAKGNYGLLDLIQALRWTSENIGFFGGDPLRITVFGSGAGGSCVNLL 290
M. musculus LGVLGFLSTGD-QAAKGNYGLLDLIQALRWTSENIGFFGGDPLRITVFGSGAGGSCVNLL 290

B


C. elegans GHE----------TNIARILVIPRGVKGFGFILRGAKHVAMP 459
H. sapiens, isoform X4 FGETRADRSKKLFRHYTVGSYDSFDTSSDCIIEEKTVVLQKKDNEGFGFVLRGAKADTPI 80
H. sapiens, isoform X2 QAETRADRSKKLFRHYTVGSYDSFDTSSDCIIEEKTVVLQKKDNEGFGFVLRGAKADTPI 650
D. rerio RLETREDRNKRLFRHYTVGSYDNFTSYSDYIIEEKNAMLQKKENEGFGFVLRGAKAETPI 542
M. musculus RHETREDRTKRLFRHYTVGSYDSLTSHSDYVIDDKVAILQKRDHEGFGFVLRGAKAETPI 669
B. taurus RHETREDRTKRLFRHYTVGSYDSLSAHSDYVIDDKVAVLQKRDHEGFGFVLRGAKAETPI 685
M. mulatta RPETREDRTKRLFRHYTVGSYDSLTSHSDYVIDDKVAVLQKRDHEGFGFVLRGAKAETPI 595
C. elegans LNFEPTAQVPALQFFEGVDMSGMAVRAGLRPGDYLLEIDGIDVRRCSHDEVVEFIQQAGD 519
H. sapiens, isoform X4 EEFTPTPAFPALQYLESVDEGGVAWQAGLRTGDFLIEVNNENVVKVGHRQVVNMIRQGGN 140
H. sapiens, isoform X2 EEFTPTPAFPALQYLESVDEGGVAWQAGLRTGDFLIEVNNENVVKVGHRQVVNMIRQGGN 710
D. rerio EEFTPTPAFPALQYLESVDVEGVAWRAGLRTGDFLIEVNGVNVVKVGHKQVVSLIRQGGN 602
M. musculus EEFTPTPAFPALQYLESVDVEGVAWRAGLRTGDFLIEVNGVNVVKVGHKQVVGLIRQGGN 729
B. taurus EEFTPTPAFPALQYLESVDVEGVAWRAGLRTGDFLIEVNGVNVVKVGHKQVVALIRQGGN 745
M. mulatta EEFTPTPAFPALQYLESVDVEGVAWRAGLRTGDFLIEVNGVNVVKVGHKQVVALIRQGGN 655
*

C


C. elegans MPTPPGQSSSQPLPAKELGYFKKIVKSYEQKQYKAGLKFAQKILTSPGFAEHGETLAMKG 60
D. rerio -------MPSITLPPKENALFKRILRCYEHKQYRNGLKFCKQILSNSKFAEHGETLAMKG 53
G. Gallus domesticus -------MPSVSLPPKENALFKRILRCYEHKQYRNGLKFCKQILSNPKFAEHGETLAMKG 53
M. musculus -------MPAVSLPPKENALFKRILRCYEHKQYRNGLKFCKQILSNPKFAEHGETLAMKG 53
H. sapiens -------MPAVSLPPKENALFKRILRCYEHKQYRNGLKFCKQILSNPKFAEHGETLAMKG 53
B. taurus -------MPAVSLPPKENALFKRILRCYEHKQYRNGLKFCKQILSNPKFAEHGETLAMKG 53
C. elegans LILNCMGKLQEAQDCVRRGLVSDLRSYVCWHVFGLIQKTEKKYDEAIKAYKRALMLEKDN 120
D. rerio LTLNCLGRKEEAYDLVRRGLRNDLKSHVCWHVYGLLQRSDKKYDEAIKCYRNALKWDKDN 113
G. Gallus domesticus LTLNCLGKKEEAYELVRRGLRNDLKSHVCWHVYGLLQRSDKKYDEAIKCYRNALKWDKDN 113
M. musculus LTLNCLGKKEEAYELVRRGLRNDLKSHVCWHVYGLLQRSDKKYDEAIKCYRNALKWDKDN 113
H. sapiens LTLNCLGKKEEAYELVRRGLRNDLKSHVCWHVYGLLQRSDKKYDEAIKCYRNALKWDKDN 113
B. taurus LTLNCLGKKEEAYELVRRGLRNDLKSHVCWHVYGLLQRSDKKYDEAIKCYRNALKWDKDN 113

D

E

Figure 4: Examples of highly-conserved and moderately-conserved regions in ASD-implicated genes across several animal species. Regions considered to be highly-conserved (boxed in black) or moderately-conserved (boxed in red) were sequences that strictly preserved the exact amino acid sequence as other spatially equivalent genes from other species, or differed by only a few amino acid substitutions that resulted in similar characteristics, respectively. Regions that showed significant conservation but differed significantly in one or more groups were boxed in green. Because the availability of equivalent orthologs for ASD implicated genes varied by species, multiple species and their ortholog sequences were considered. Each letter of the sequence represents the one-letter code for the twenty known amino acids, and colors represent different properties of the amino acid residue: Red denoting small, hydrophobic or aromatic residues, blue for acidic residues, purple for basic residues, and green for residues containing hydroxyl, sulfhydryl or amino groups. Asterisks at the bottom of each grouped row of multiple aligned sequences indicate positions with a single conserved residue. Colon symbols indicate conservation between groups of strongly similar properties; roughly equivalent to scoring > 0.5 in the Gonnet PAM 250 matrix. A period indicates conservation between groups of weakly similar properties as below; roughly equivalent to scoring =< 0.5 and > 0 in the Gonnet PAM 250 matrix. Numbers to the right of each line of peptide sequence indicate the number of the starting amino acid of that line that it corresponds to in the overall numbered gene sequence. **(A)** MSA example of the human (*H. sapiens*) gene NGLN1 and its ortholog in *C. elegans* (worms), *D. Melanogaster* (fruit flies), *R. norvegicus* (rats), and *M. Musculus* (mice). **(B)** MSA example of the human gene SHANK3 (two possible isoforms) and its ortholog in *C. elegans*, *D. rerio* (zebrafish), *M. musculus*, *B. taurus* (cattle) and *M. mulatta* (rhesus monkeys). **(C)** MSA example of the human gene NAA15 and its ortholog in *C. elegans*, *D. rerio*, *G. gallus domesticus* (chickens), *M. musculus*, and *B. taurus*. **(D)** MSA example of the human gene DYRK1A and its ortholog in *D. melanogaster*, *C. elegans*, *D. rerio*, *B. taurus* and *M. musculus*. **(E)** MSA example of human gene SETD5 and its ortholog in *C. elegans*, *M. musculus*, *G. gallus domesticus*, *D. rerio* and *O. niloticus* (tilapia).

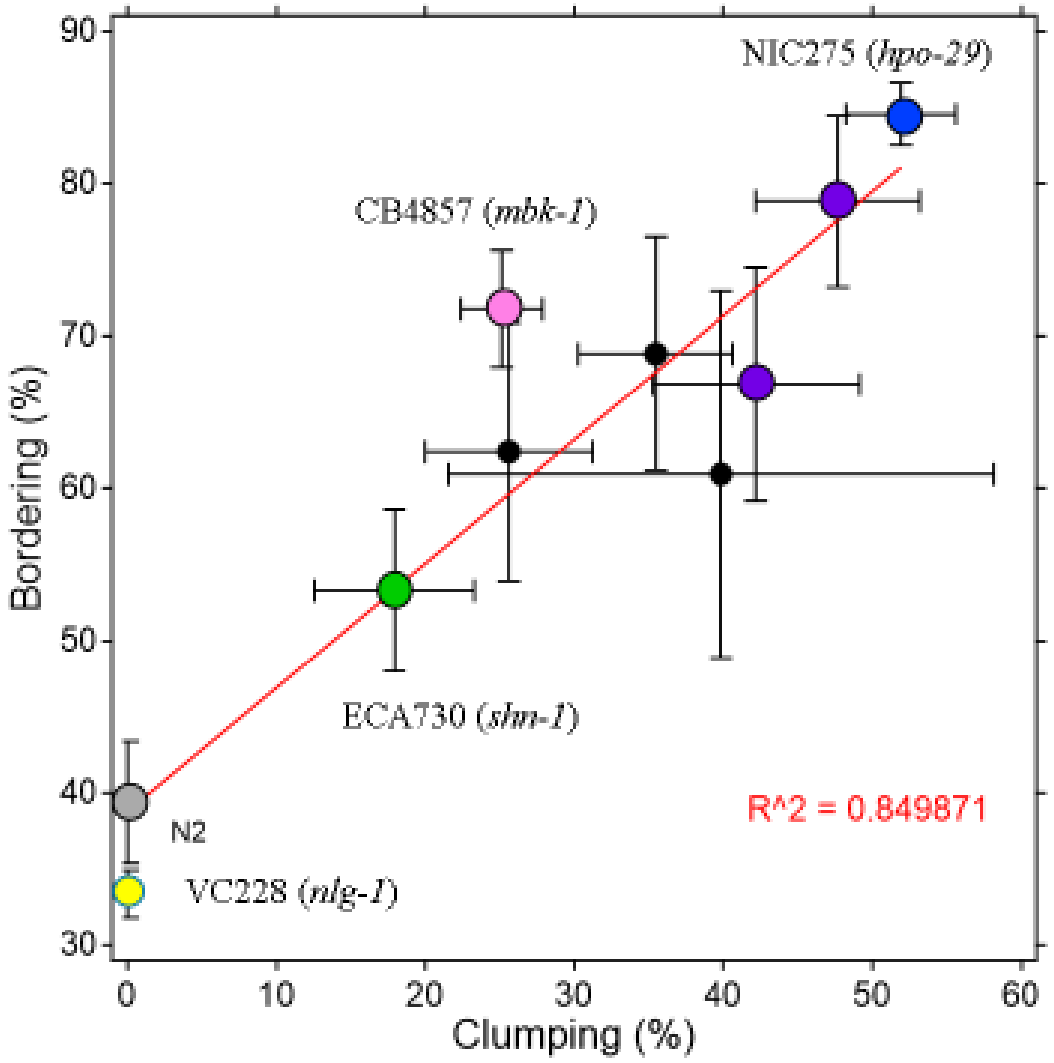


Figure 5: Bordering and clumping social behavior are highly correlated but not synonymous. Using the calculation for Pearson correlation coefficient, the coefficient of determination was determined to be 0.849871, suggesting the relationship between bordering and clumping behavior is well fit to the regression model. Color-filled graphical points are to denote worm strains with distinct social behavior: (From left-to-right on the graph) Strain VC228 (*nlg*-1) is denoted in yellow, N2 as grey, ECA730 (*shn-1*) as green, CB4857 (*mbk-1*) as pink, positive social controls RC301 and CB4932 as two purple points, and then farthest right is strain NIC275 (*hpo-29*), which is denoted in blue.

DISCUSSION

Social behavior assays

Evaluation of bordering and clumping behavior in ASD strains displayed that strains with variants in ASD candidate genes NLGN-1, SHANK3, DYRK1A, and SETD5 displayed significant defects in social behavior compared to controls and other wild-isolate strains (Figure 5). Strains VC228 (nlg-1) and ECA730 (*shn-1*) both display a significant decrease in bordering, with ECA730 worms also decreasing in clumping behavior compared to other wild-isolate strains. CB4857 (*mbk-1*) worms demonstrate average bordering behavior as expected of wild-isolate strains, but shows a significant decrease in clumping behavior ($p<0.005$). Interestingly, strain NIC275 (*mbk-1*) worms consistently show a slight increase in both social feeding behaviors compared to other wild-isolate strains and positive controls.

MSA analysis

MSA analysis reveals multiple areas of high conservation for each ASD gene studied, with the exception of SETD5. Although different animal species were used for each MSA, human genes and their ortholog counterparts showed significant variability in amino acid identity, but conserved the order of residue chemical properties (denoted by sequence color scheme) relatively well, most likely to preserve the general structure of the gene product even though there are slight sequence modifications. Particularly of interest, human gene NAA15 was almost completely well-conserved, with each ortholog counterpart almost being identical in amino acid length. This suggests that almost all of the information encoded by this gene is likely sensitive to mutations, and requires more analysis. Additional identification of highly-conserved areas for the entire amino acid sequence for each gene is provided in the supplemental figures for this project.

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK

Our results demonstrate that strains with predicted change-of-function mutations in the orthologs of human genes DYRK1A, SHANK3, NLGN1 and SETD5 demonstrated significantly lower levels of social behavior compared to other social strains with normal sequence of these genes, and that clumping and bordering social behavior is highly correlated but not synonymous. Identifying causal genes and their distinct variants for any strong ASD phenotype is imperative for understanding the biological foundations of this heterogeneous disorder.

Although the bordering and clumping results generally suggest consistent measures of bordering and clumping behavior in each strain, only eight trials per strain is still relatively low. Future plans for this study include repeating bordering and clumping assays for strains that demonstrated high variability in the average for each social behavior, particularly strain JU3282 (*hpo-29*, standard deviation of 24.06 and 36.55, respectively). By completing more trials for each strain, we hope to tighten the parameters of the observed bordering and clumping averages to draw more conclusive causal relationships between the tested variants and changes in social behavior. Additionally, due to the seemingly promising results of the VC288 (N2 background) strain (33.94% bordering, 0% clumping), more social behavior assays should be performed on nlg-1 strains with the same variant but with a wild-isolate background. Because this particular strain does not contain the N2 genetic background that the other wild-isolates have, it is difficult to conclude whether the defects in social behavior are due to the nlg-1 variant or other confounding background genetics. However, VC228 worms showed significantly lower bordering averages compared to control N2 worms (33.94%, $p=0.12$), which may suggest that the nlg-1 variant is *increasing* the resistance to bordering already normally seen in N2 worms. No claim can be made about the nlg-1 variant potentially reducing clumping as well, because both VC228 and N2 worms demonstrated 0% clumping.

Continuation of this study should also include expansion of genes and variants considered for social defects. Each gene encodes long and complex sequences of nucleotides, encoding rich genetic information that may produce a valuable protein, modify gene expression, or contribute to other fundamental cellular processes. Because genes encode so much valuable information, it is prone to many potential mutations or variants that may dramatically affect the structure and/or function of that gene product. To make this experiment feasible, only one to two variants could be studied for each of the available human gene-orthologs found in *C. elegans*. Although this study may lay the groundwork for studying particular variants in a narrowed down list of Category 1 ASD genes, further investigation is needed to determine the effects of other possible variants in these genes.
Once ASD genes and variants are confirmed to be associated with *C. elegans* social defects, CRISPR gene-editing will help clarify the causal nature of these particular mutations with conserved population-wide variability in *C. elegans* social behavior.

Designing reagents to CRISPR target candidate genes and performing a full knock-in mutation will conclusively determine that particular variant is solely responsible for the social defect, rather than some other unknown variant or unidentified gene interaction that may be present in these wild-isolate strains.

Additionally, there has been recent speculation about bordering and clumping being considered more as an oxygen avoidance behavior in worms rather than a social behavior. *C. elegans* are known to be highly sensitive to their environmental levels of oxygen due to the uneven distribution of air in their natural soil and water environments, presenting the constant risk of hypoxia or hyperoxia. In a study done by Chang et al. in 2006, hyperoxia avoidance behavior was attributed to the activity oxygen-sensing neurons that were mediated by *npr-1*. To determine whether decreases in bordering and clumping behavior are due to differences in oxygen sensing dysfunction or because of social behavior defects, repetition of trials in an oxygen chamber with consistent air distribution will confirm whether our variants of interest are truly causal for social defects rather than increasing oxygen avoidance behaviors.

In summary, understanding the genetic mechanisms that underlie the onset of ASD phenotypes will modernize how this disorder and other similar intellectual and neurodevelopmental disorders are understood and treated. Establishment of what regions are considered highly conserved in human genes and their orthologous counterparts will aid future geneticists and researchers to predict and recognize significant variants in ASD candidate genes.

ACKNOWLEDGEMENTS

Endless gratitude to my stellar advisor, Dr. Jonathan Pierce. His patience, gentle but firm mentoring, and contagious enthusiasm for science has been imperative not only for the completion of this research, but for the progression of my entire academic career as an undergraduate. He has introduced me to countless opportunities to expand my interests in intellectual disabilities beyond the lab setting and has added so much richness and depth to my undergraduate experience. I am eternally grateful for him. Thank you to our lab manager and hero Susan Rozmiarek, graduate student Benjamin Clites, and fellow undergraduates Courtney Williams, Sofia Valdivies-co-Sinyakov, and Abigail Tyree for aiding in the data collection and completion of this thesis.

CITATIONS

Calahorro, Fernando, and Manuel Ruiz-Rubio. "Caenorhabditis Elegans as an Experimental Tool for the Study of Complex Neurological Diseases: Parkinson's Disease, Alzheimer's Disease and Autism Spectrum Disorder." SpringerLink, Springer-Verlag, 8 Nov. 2011, link.springer.com/article/10.1007/s10158-011-0126-1.

Deweerdt, Sarah. "Roundworm Roundup May Reveal Function of Autism Genes." Spectrum, 6 Nov. 2018, www.spectrumnews.org/news/roundworm-roundup-may-reveal-function-autism-genes/.

"Engineering Bacteria to Form a Biofilm and Induce Clumping in Caenorhabditis Elegans." ACS Synthetic Biology, pubs.acs.org/doi/abs/10.1021/sb4001883.

"GO Term: Peptide Alpha-N-Acetyltransferase Activity." N, www.candidagenome.org/cgi-bin/GO/go.pl?goid=4596.

Hu, Zhitao, et al. "Neurexin and Neuroligin Mediate Retrograde Synaptic Inhibition in C. Elegans." Science, American Association for the Advancement of Science, 24 Aug. 2012, science.sciencemag.org/content/337/6097/980.

Hunter, Jerrod W., et al. "Neuroligin-Deficient Mutants of C. Elegans Have Sensory Processing Deficits and Are Hypersensitive to Oxidative Stress and Mercury Toxicity." Disease Models & Mechanisms, The Company of Biologists Ltd, 1 May 2010, dmm.biologists.org/content/3/5-6/366.

Kim, Woojin, et al. "OrthoList 2: A New Comparative Genomic Analysis of Human and Caenorhabditis Elegans Genes." *Genetics*, Genetics, 1 Oct. 2018, www.genetics.org/content/210/2/445.

"Leo Kanner's 1943 Paper on Autism." Spectrum, 2 Apr. 2019, www.spectrumnews.org/opinion/viewpoint/leo-kanners-1943-paper-on-autism/.

Macosko, Evan Z., et al. "A Hub-and-Spoke Circuit Drives Pheromone Attraction and Social Behaviour in C. Elegans." Nature News, Nature Publishing Group, 6 Apr. 2009, www.nature.com/articles/nature07886.

Mossa, Adele, et al. "SHANK Genes in Autism: Defining Therapeutic Targets." Progress in Neuro-Psychopharmacology and Biological Psychiatry, Elsevier, 22 Nov. 2017, www.sciencedirect.com/science/article/pii/S0278584617307522.

"Natural Variation in a Neuropeptide Y Receptor Homolog Modifies Social Behavior and Food Response in C. Elegans." Cell, Cell Press, 11 Apr. 2001, www.sciencedirect.com/science/article/pii/S0092867400816098.

Reddy, Kirthi C., et al. "Caenorhabditis Elegans NPR-1–Mediated Behaviors Are Suppressed in the Presence of Mucoid Bacteria." PNAS, National Academy of Sciences, 2 Aug. 2011, www.pnas.org/content/108/31/12887.

Satterstrom, F Kyle, et al. "Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism." *Cell*, U.S. National Library of Medicine, 6 Feb. 2020, www.ncbi.nlm.nih.gov/pubmed/31981491.

"Strain Information." *Strain: VC228, Genotype: Nlg-1(ok259) X. - Caenorhabditis Genetics Center (CGC) - College of Biological Sciences*, cgc.umn.edu/strain/VC228.

Tong XJ et al., "Retrograde Synaptic Inhibition Is Mediated by α-Neurexin Binding to the α2δSubunits of N-Type Calcium Channels." Neuron. 2017 Jul 19;95(2):326-340.e5.

Wong WR, Brugman KI, Maher S, Oh JY, Howe K, Kato M, Sternberg PW. Autism-associated missense genetic variants impact locomotion and neurodevelopment in Caenorhabditis elegans. Hum Mol Genet. 2019 Jul 1;28(13):2271-2281. doi: 10.1093/hmg/ddz051. PMID: 31220273; PMCID: PMC6586145.

Discovery of Inhibitory Biomolecules of D-Alanine: D-Alanine Ligase of *Staphylococcus aureus* Using High-Throughput Virtual Screening

Kaleigh Copenhaver
The University of Texas at Austin
Texas Undergraduate Research Journal

ABSTRACT

New antibiotic-resistant strains of *Staphylococcus aureus* continue to emerge, creating additional challenges to not only treating symptoms of infections but also preventing the spread of the bacteria. D-alanine: D-alanine Ligase (DDl) is an essential enzyme of S. aureus that assists in peptidoglycan synthesis[1]. Current antibiotics, such as vancomycin, target and disrupt this process, but antibiotic resistance limits the use of this treatment[2]. The objective of this research is to discover novel ligands via virtual screening that can bind and potentially inhibit DDl. Computational ligand libraries of varying sizes and molecular scaffolds were docked against the DDl crystal structure using two virtual screening softwares, GOLD and ICM. A total of 139,091 computational ligands were virtually docked, each ligand was given an overall binding score derived from the intermolecular forces with the protein, and rank-ordered. Thus far, virtual screening has provided library docking scores ranging from 59.96 to 92.22 and 15.86 to -34.72 for GOLD and ICM docked compounds, respectively. In the next phase of the project, the enzymatic activity of DDl and inhibition abilities of novel compounds will be measured through biochemical assays to determine the ligand's viability as a potential pharmaceutical.

INTRODUCTION

Staphylococcus aureus is the leading cause of infection in U.S. healthcare facilities and an estimated 15 to 30% of all individuals carry the bacteria on the surface of their skin or mucous membranes[3]. These rates jump to around 80% for healthcare workers, immunocompromised individuals, and hospitalized patients[4,5]. These carriers are not only at a higher risk of being infected themselves, but also assisting in the spread of the bacteria[6]. Infections typically result in symptoms of skin rashes, redness, or irritation, accompanied with a fever, though a majority of individuals never experience symptoms[7]. If the bacteria enters the bloodstream through direct contact with an infected individual or contaminated surface, a plethora of additional infections and ailments can occur, such as pneumonia, endocarditis, sepsis, and in severe cases, death, if the infection is left untreated or if underlying diseases are present[7,8,9]. In 2017, over 20,000 of the 119,000 people in the United States with a bloodstream S. aureus infection died[5].

D-alanine: D-alanine ligase (DDl) is an essential enzyme of S. aureus that assists in peptidoglycan production[4]. Peptidoglycan is vital to a bacterial cell wall as it provides structural support for the cell to withstand the osmotic pressure gradient that exists between the cytoplasm and materials outside of the cell[4]. DDl assists in the formation of cross-linkages between the N-acetylmuramic acid (NAM) and N-acetylglucosamine (NAG) polymers in the peptidoglycan layer[2]. Disruption of peptidoglycan cross-linkage formation can lead to a weakened

cell-wall, and eventually, bacterial death[2]. Inhibition of the enzyme by a novel ligand, such as ZINC_1 shown in Fig. 6, has the potential to disrupt cell wall stability and lead to cell lysis[10].

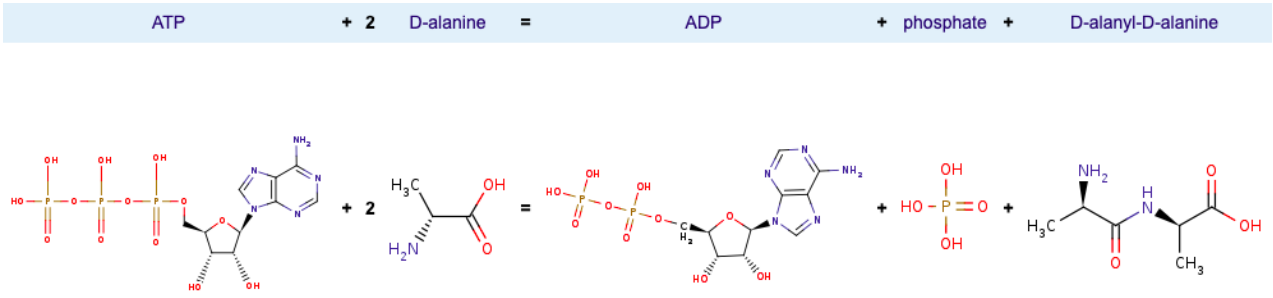


Figure 1: *D-alanine:* D-alanine Ligase reaction mechanism obtained from BRENDA (E.C.: 6.3.2.4)

Additionally, antibiotic resistant strains, such as MRSA (Methicillin-Resistant S. aureus), have rapidly expanded beyond the previous confines of high-risk hospital settings, which have made traditional antibiotics an ineffective treatment method[1,11]. MRSA infections in the United States were declining 17% annually from 2005 to 2013; however, this progress has slowed in recent years as antibiotic resistance to medications like penicillin, methicillin, and tetracyclines continues to pose a threat and new strains of S. aureus emerge[5,6]. Traditional oral medicines are being replaced by more intensive intravenous medications with increased side effects and risks as the bacteria continues to adapt to treatment methods[2]. Vancomycin is an intravenous antibiotic used to treat S. aureus infections; however, as resistance to other medications continue to increase as well as emergence of Vancomycin resistant bacteria, its use is limited to severe cases[2]. This antibiotic targets the cell wall synthesis of gram-positive bacteria by binding to the D-alanyl: D-alanine termini of the peptidoglycan layer to disrupt nearby cross linkage formation and ultimately lead to cell death[2,11]. Virtual screening is often used in drug development research to dock computational ligands against a target protein to determine a compounds' binding potential. This method allows large numbers of compounds to be tested quickly and in a cost-effective manner. The screening software scans through various binding conformations for each ligand and estimates where the ligand would be placed in the active site of the protein. Each ligand is assigned an overall binding score derived from the intermolecular forces with the protein[12,13]. This score is an indication of the binding ability for a given ligand, and top-scoring ligands can later be tested in vitro using biochemical assays. While virtual screening cannot definitively determine a compound's ability to inhibit DDl, virtual screening does significantly narrow the field of ligands to be tested in the lab. The inhibition of DDl and, in turn, peptidoglycan production has effectively prevented cell growth in the past, and indicates that novel compounds found through virtual screening could inhibit DDl in a similar manner. This work intends to discover novel ligands via virtual screening that have strong predicted binding abilities against DDl and have the potential to be made into new pharmaceuticals to treat S. aureus.

METHODS

Virtual Screening

Virtual screening was completed through the Edu POD of the Biomedical Research Computing Facility (BCRF) that has three compute servers each with two core, 52-thread CPU, and 1 TB of RAM. The EDU POD has a shared 96 TB of storage for the computer servers.

Two virtual screening softwares were used to determine binding abilities of novel compounds. The first software, Genetic Optimisation for Ligand Docking (GOLD), uses a genetic algorithm that docks ligands into a specified protein binding site[12]. GOLD scans through various ligand conformations, based on their rotatable bonds, and places each compound in the active site of the protein. Each ligand is assigned a Fitness Score where higher scores are associated with stronger binding abilities and therefore greater predicted drug activity[12]. This score is a summation of changes in free energies of intermolecular forces associated with binding, such as Van der Waals forces, hydrophobic interactions, Hydrogen bonds, and steric strains (see equation below)[12]. GOLD provides a sub-score for each of these categories.

$$\Delta G_{bind} = \Delta G_{vdw} + \Delta G_{H-bond} + \Delta G_{hydrophobic} + \Delta G_{rotor} + \Delta G_0$$

The second software, Internal Coordinate Mechanics (ICM), uses a global optimization algorithm to dock ligands13. The ICM score is dependent on the Van der Waals interactions, Hydrogen bond interactions, free energy changes associated ligand binding and hydrophobic interactions, desolvation energy of Hydrogen bond donors and acceptors, solvation electrostatic energy of ligand binding, and a size correction term (see equation below)13. Unlike GOLD scores, smaller ICM scores predict better binding abilities.

$$\Delta G = \Delta E_{\mathrm{IntFF}} + T\Delta S_{\mathrm{Tor}} + \alpha_1 \Delta E_{\mathrm{HBond}} + \alpha_2 \Delta E_{\mathrm{HBDesol}} +$$

$$\alpha_3 \Delta E_{\mathrm{SolEl}} + \alpha_4 \Delta E_{\mathrm{HPhob}} + \alpha_5 Q_{\mathrm{Size}}$$

Identifying the Protein Structure

Before computational compounds can be screened, an X-ray crystallography protein structure was identified and analyzed. A Vancomycin-resistant Staphylococcus aureus (VRSA) D-Alanine: D-Alanine ligase crystal structure was obtained from the PDB website (PDB ID: 3N8D) with a resolution of 2.30 Å14. and steric strains (see equation below)[12]. GOLD provides a sub-score for each of these categories. Higher resolutions provide more confidence that the locations of atoms in space are actually where the PDB says they are. The second repeating chain of the dimer was removed in PyMOL. The protein structure contained one Phosphoaminophosphonic acid-adenylate ester (ANP) molecule as well as a chloride ion and a sulfate ion. ANP is an analog of ATP and can therefore predict what the protein structure may look like in complex with ATP[15]. The sulfate and chloride ions remained from the crystallography procedure and were removed before proceeding with virtual screening.

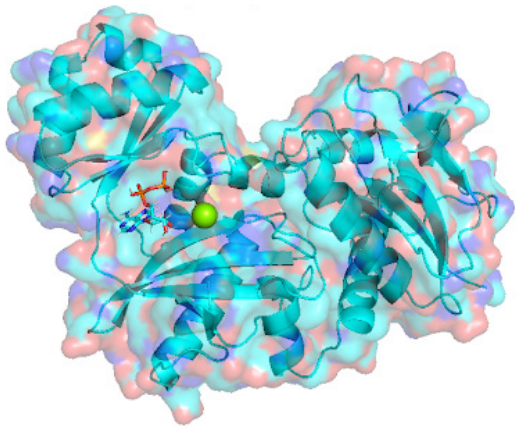


Figure 3: X-ray crystallography image of D Alanine: D-Alanine Ligase of Staphyloccocus aureus (PDB: 3N8D). The protein is shown with a transparent surface and ribbon structure with carbons colored blue. ATP is shown as sticks in the active site with carbons colored blue. The green sphere is a magnesium cofactor.

Next, the updated protein was loaded to the MolProbity website where Hydrogens were added and any asparagine, glutamine, and histidine flips were recorded. The flips allow for a more precise estimation of Hydrogen bonding and steric interactions that allows MolProbity to better predict potential clashes[16]. None of the recorded flips occurred in the active site and were therefore not an issue when docking. The protein structure without Hydrogens was downloaded to be used during docking, wherein the Hydrogens would be added back by the molecular docking program (GOLD or ICM). Overall, MolProbity's structure analysis did not show any major clashes or strains that would impact the docking results and yielded an overall good score for the quality of this structure.

XQuartz was used to access the GOLD and ICM graphical user interfaces. Hydrogens were added to the protein, waters removed, and the original ANP ligand was extracted. The binding site was defined at the D-Alanine docking site where the ANP ligand was located.

Ten positive controls expected to bind relatively well were chosen from The Binding Database or previous literature predicting these ligands would bind well. Additionally, the original ANP ligand was used as a validation dock to determine if the GOLD and ICM softwares were able to dock the ligand in the correct location. Next, five random negative control ligands were chosen and the physio-chemical properties, such as molecular weight, Hydrogen bond donors and acceptors, and LogP, for all of the control ligands was recorded. Each control ligand could not violate more than one of Lipinski's Rule of Five to determine the molecule's predicted drug-likeness. Each ligand file was uploaded to OpenBabel in order to protonate each molecule at physiological pH and generate a three-dimensional, low energy conformation. After the control ligands were chosen and prepared, the ligand files were concatenated into one file for docking.

For the GOLD runs, control docking was completed using one processor and an autoscale of 2.0. ICM does not use an autoscale like GOLD, but the output of docking results can be altered by adjusting the threshold. This is more important for the library docking and a large threshold was used to achieve scores for all of the controls. The output ligands were analyzed in PyMOL to determine polar contacts and verify poses.

Library Docking

A total of 139,091 experimental ligands from two small libraries, In House Compounds (IHC) and CB306, and two large libraries, HitFinder9 (HF9) and ZINC, were virtually docked against the DDl protein. The IHC library was the smallest of the four libraries screened and consisted of 40 ligands from compounds already owned by the Virtual Drug Screening lab. The CB306 library contained 306 ligands, as the name suggests, from ChemBridge[17]. The HF9 library contained 14,400 ligands from Maybridge's HitFinder compound library[18]. The ZINC library was the largest library screened with

All-Atom Contacts	Clashscore, all atoms:	6.67		88th percentile* (N=1784, all resolutions)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	18	6.23%	Goal: <0.3%
	Favored rotamers	254	87.89%	Goal: >98%
	Ramachandran outliers	1	0.31%	Goal: <0.05%
	Ramachandran favored	309	96.26%	Goal: >98%
	Rama distribution Z-score	-1.56 ± 0.39		Goal: abs(Z score) < 2
	MolProbity score^	2.22		63rd percentile* (N=27675, 0Å - 99Å)
	Cβ deviations >0.25Å	1	0.32%	Goal: 0
	Bad bonds:	4 / 2702	0.15%	Goal: 0%
	Bad angles:	7 / 3661	0.19%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	2 / 13	15.38%	Expected: ≤1 per chain, or ≤5%
Low-resolution Criteria	CaBLAM outliers	4	1.3%	Goal: <1.0%
	CA Geometry outliers	1	0.32%	Goal: <0.5%
Additional validations	Chiral volume outliers	0/409		
	Waters with clashes	3/134	2.24%	See UnDowser table for details

Figure 2: MolProbity analysis of DDl X-ray crystallography structure (PDB: 3N8D)[15].

a total of 124,271 compounds from the ZINC database[10].

The small libraries screened using GOLD's docking software were completed in a similar manner to the control libraries using one processor and an autoscale of 2.0. The large libraries were also screened using an autoscale of 2.0, but were split between six processors to complete the docking process in a reasonable amount of time. The top 10% of ligands from the large libraries were saved for analysis. For ICM docking, the threshold was adjusted for the various libraries to save only the top scoring ligands from each library. The ZINC library was screened with a threshold of -32 and the other three libraries were screened with a threshold of -25. Only ligands with scores below this threshold are provided by ICM. A larger magnitude for the threshold leads to more selective screening where only compounds with scores equal to or below than this threshold were saved. ZINC was the only library that contained ligands with scores below the -32 threshold, so the remaining libraries were docked with a larger threshold in order to generate results for analysis.

RESULTS AND DISCUSSION

Control Results

Overall, control ligands received docking scores from both GOLD and ICM that validated the virtual screening softwares and protein structure for further screening with experimental compound libraries. For the most part, control docking with GOLD resulted in mid-to high-scoring positive controls and low-to mid-scoring negative controls, as expected. GOLD scores ligands on a positive scale where ligands scoring above 75-80 could be considered to bind relatively well. The negative controls were chosen at random and while some received a score of 0 and were not placed in the active site at all, others, like N- (4-methyl-2-pyridinyl)-2-(1-naphthyl)acetamide, resulted in relatively high scores at 63.54 and 64.31 for differing conformations.

ICM control docking had similar results, but some of the positive controls scored above the anticipated threshold. The positive control ligands were chosen based on their known ability to inhibit DDl or

Chemical Name	GOLD Score	ICM Score	S(PLP)	S(hbond)	Ligand Efficiency	MW	LogP	H-Bond Donor	H-Bond Acceptor
9-Acridinylamine	71.20	-6.11	-66.60	1.59	0.154	463	5.0	2	6
ADP	68.24	-21.80	-51.62	5.43	0.160	427	-4.6	6	14
9-Acridinylamine	67.55	-2.27	-62.26	2.00	0.146	463	5.0	2	6
ZINC4416106	65.61	-3.24	-61.01	2.00	0.170	386	5.6	1	4
ZINC4416106	65.36	-1.47	-54.87	3.94	0.169	386	5.6	1	4
N-[(4-Hydroxypyrrolidin-3-yl)methyl]-2-[4-(thiophen-2-ylsulfonylamino)phenyl]acetamide	64.96	-1.83	-55.74	3.76	0.164	396	1.2	4	7
ANP	64.86	-16.44	-45.42	6.60	0.128	506	-6.0	8	17
N-(4-methyl-2-pyridinyl)-2-(1-naphthyl)acetamide	64.31	-18.28	-62.38	0.94	0.233	276	3.6	1	2
N-(4-(2-oxo-2-(pyrrolidin-1-yl)ethyl)phenyl)thiophene-2-sulfonamide	63.86	-7.25	-57.34	1.61	0.182	351	2.3	1	5
N-(4-(2-oxo-2-(pyrrolidin-1-yl)ethyl)phenyl)thiophene-2-sulfonamide	63.75	-7.35	-57.69	1.45	0.182	351	2.3	1	5
N-(4-methyl-2-pyridinyl)-2-(1-naphthyl)acetamide	63.54	-17.36	-62.25	0.00	0.230	276	3.6	1	2
9-Methoxyellipticine	63.28	-11.34	-55.14	1.01	0.190	333	3.6	1	1
9-Methoxyellipticine	62.61	-9.77	-51.63	2.00	0.188	333	3.6	1	1
AKOS017196493	61.84	-1.74	-56.03	2.53	0.168	369	1.5	2	6
N-[(4-Hydroxypyrrolidin-3-yl)methyl]-2-[4-(thiophen-2-ylsulfonylamino)phenyl]acetamide	61.05	-1.45	-51.94	3.47	0.154	396	1.2	4	7
AKOS017196493	60.75	-1.93	-58.16	1.4	0.165	369	1.5	2	6
ADP	60.06	-14.33	-35.98	7.18	0.141	427	-4.6	6	14
Elliptinium	56.55	-14.19	-54.73	0.61	0.204	277	4.6	2	1
Elliptinium	55.25	-14.08	-53.04	0.74	0.199	277	4.6	2	1
MFCD01833137	53.06	-12.41	-51.09	0.91	0.162	327	3.3	2	3
MFCD01833137	51.35	-11.33	-49.73	0.00	0.157	327	3.3	2	3
3-Chloro-2,2-dimethyl-N-[4-(trifluoromethyl)phenyl]propanamide	42.72	-9.14	-38.37	1.69	0.153	280	3.4	1	4
N-(4-ethylphenyl)-2,2,2-trifluoroacetamide	42.72	-6.80	-41.23	1.00	0.197	217	3.3	1	4
3-Chloro-2,2-dimethyl-N-[4-(trifluoromethyl)phenyl]propanamide	42.53	-9.22	-39.01	1.70	0.152	280	3.4	1	4
N-(4-ethylphenyl)-2,2,2-trifluoroacetamide	41.93	-11.13	-39.32	0.99	0.193	217	3.3	1	4
Acetamide	41.11		-36.15	1.85	0.246	167	2.1	1	2
2,2,2-trifluoro-N-(4-methylphenyl)acetamide	40.97	-11.03	-35.62	1.92	0.202	203	2.6	1	4
Acetamide	40.44		-35.57	1.83	0.242	167	2.1	1	2
2,2,2-trifluoro-N-(4-methylphenyl)acetamide	40.29	-10.71	-35.16	1.92	0.198	203	2.6	1	4
D-cycloserine	28.07	-5.47	-21.87	2.07	0.275	102	-1.5	2	3
D-cycloserine	25.80	-5.45	-13.80	4.00	0.253	102	-1.5	2	3

Table 1: Control docking results ranked by GOLD score. Positive controls are highlighted in green and negative controls are highlighted in red.

similar protein structures. ICM scores ligands on a negative scale, unlike GOLD, where larger magnitudes are expected to bind more favorably. An ICM score lower than -25 would be considered to bind well. One of the top-scoring control ligands determined by GOLD, 9-acrylamide, received GOLD scores of 67.55 and 71.20 yet only had ICM scores of -2.265 and -6.111 (Table 1). We expected the positive controls screened with ICM to have more favorable (lower) scores than the observed scores.

The variations in binding predictions of the positive controls raised the question of how strong was the correlation between the overall scores between the two softwares. The ICM and GOLD scores were plotted against one another, received an R^2 value of 0.0016 showing a low correlation between the scoring data of the two software programs (Fig. S1). The control scores had less variation when screened with GOLD than with ICM (Fig. S2). The GOLD results showed a strong correlation of binding scores with an R^2 value of 0.8890, while ICM results showed an extremely low correlation of scores with an R^2 value of 0.0005 (Fig S2).

One positive control ligand, D- cycloserine (DCS), had surprisingly low scores for both GOLD and ICM screening. DCS is another common antibiotic used to treat S. aureus infections and would be expected to have good overall binding scores[19]. While GOLD and ICM predictions may be inaccurate, it is more likely that the small size of the molecule meant the screening softwares had fewer atom interactions to provide a score for. Additionally, recent evidence has shown the binding of DCS to DDl is slow and DCS may not be a reversible competitive inhibitor[20,21]. This slow-onset of binding may also explain DCS's unfavorable GOLD and ICM scores. The ligand efficiency score was used to normalize the overall scores and account for molecular weight disparities. This normalization leads to D-cycloserine being the highest scoring positive control ligand.

While the control docking for the two softwares had somewhat dissimilar results, neither resulted in scores too unusual that would invalidate either the protein structure or software for library docking use. The original X-ray crystallography DDl structure contained an ANP ligand in the active site[14]. ANP

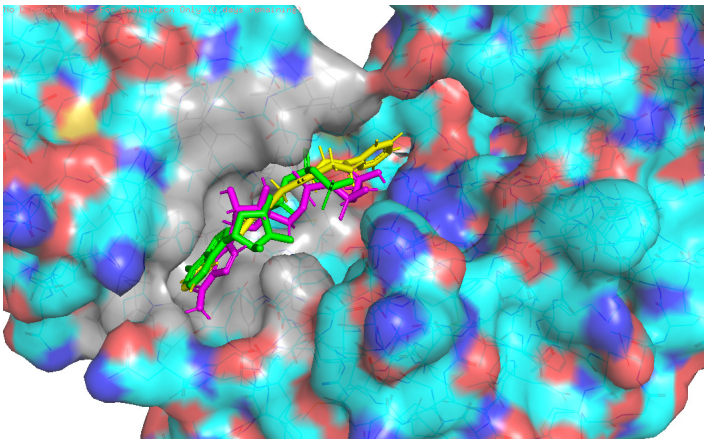


Figure 4: ANP docked in the active site of DDl (PDB: 3N8D). The ligands are shown as sticks with the original X-ray crystallography ANP ligand colored green, the GOLD docked ANP ligand colored pink, and the ICM docked ANP ligand colored yellow. The protein is shown as lines with carbons colored blue and a semi-transparent surface. The docking site is colored grey.

GOLD Screening Results

Out of the four libraries screened with GOLD, the ZINC library had the highest-scoring ligands and was the only library to achieve scores above 90. The top 5 ligands from this library scored higher than any of the other top compounds from the HF9, CB306, or IHC libraries (Table 2). The ZINC library also had the widest range of scores and some of the lowest scoring compounds, but this is likely due to its large size in comparison to the other libraries screened. IHC had the highest median GOLD score, yet the highest scoring ligand from this library, IHC_1, still scored lower than the top 5 ligands from both the ZINC and HF9 libraries.

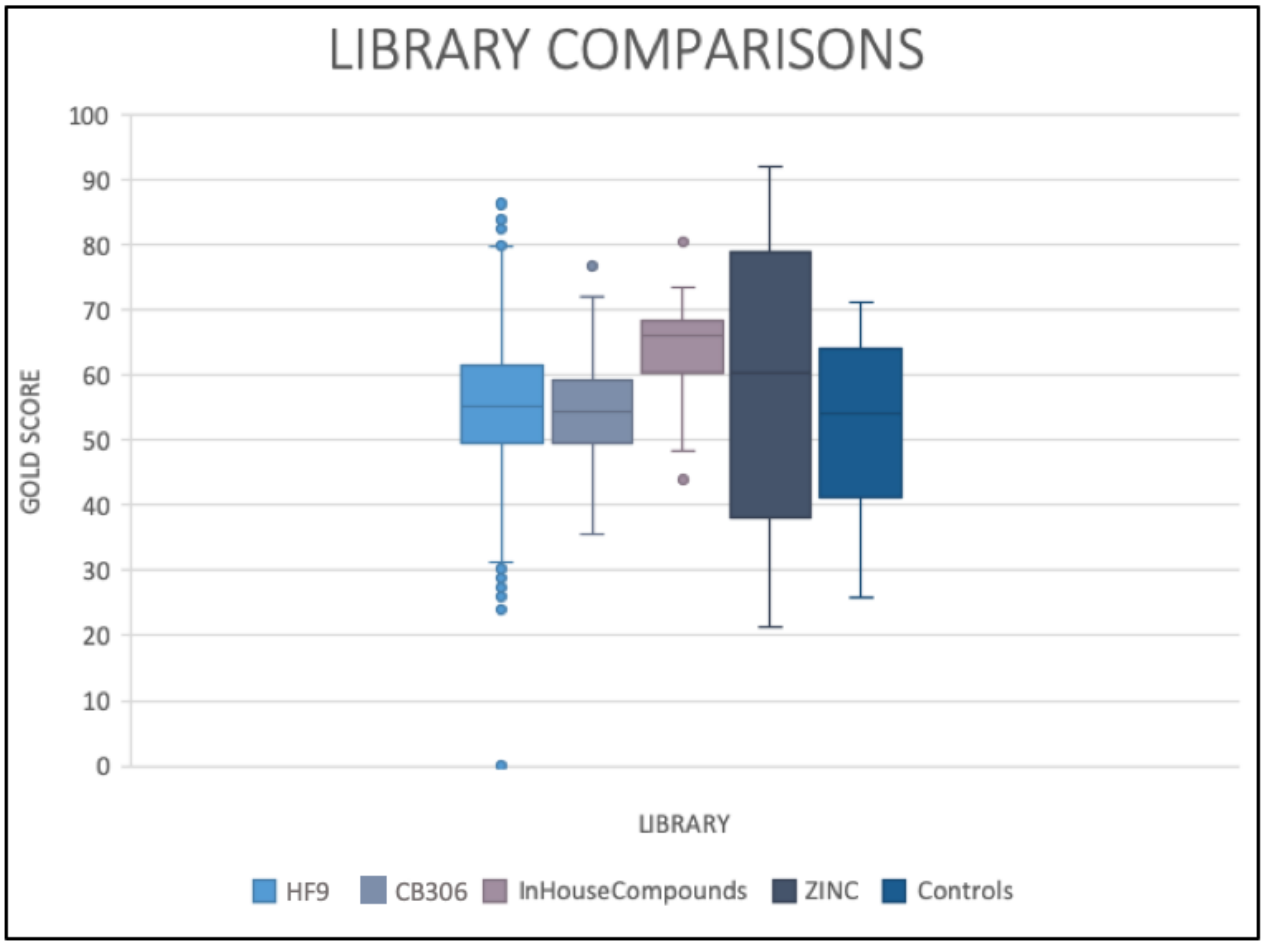


Figure 5: Boxplot comparison of GOLD scores for experimental libraries. The data is shown through a five-factor summary with the y-axis displaying the overall GOLD score. The bottom of the rectangular box represents the first quartile, the middle line is the median score, and the top of the rectangular box represents the third quartile of the data set. The minimum and maximum of the data sets are found by subtracting 1.5 multiplied by the IQR from the first quartile score and by adding 1.5 multiplied by the IQR to the third quartile score, respectively. Outliers are shown as dots. ZINC has the largest range of scores and IHC has the highest median score.

The CB306 library had some of the lowest scoring GOLD compounds in comparison to the other three libraries. Two ligands, CB306_4 and CB306_5, received overall GOLD scores of 70.56 and 69.32, respectively[17]. When normalized, CB306_4 became the top-scoring compound with a ligand efficiency score of 0.239. CB306_5's score also improved with normalization and received a ligand efficiency score of 0.203. These two compounds had "h(bond)" subscores above 3.0, which is better than all but two of the top-scoring compounds.

The "h(bond)" score is a metal correction score that calculates Hydrogen bond free energies from their actual distance as opposed to a pre-calculated distance[11]. The combination of large "h(bond)" and ligand efficiency scores indicate that these ligands may have better binding abilities in vitro than GOLD's overall score originally predicted.

The top 5 scoring compounds were all obtained from the ZINC library and their docking poses are visualized in Fig. 6. All five compounds had GOLD scores above 90. GOLD predicted binding to several specific amino acids: Lys-130, Ser-183, Ser-184, and Asn-305. Almost all of the top ZINC compounds had polar contacts to one, if not all, of these amino acids. This indicates these amino acids may be important for binding compounds to DDl. The top scoring compound, ZINC_1, had a GOLD score of 92.22 and

a ligand efficiency score of 0.209. Additionally, this molecule visually fits into the active site groove well and from GOLD's analysis would be expected to successfully bind to DDl. Surprisingly, this compound received the worst ICM score out of the top GOLD compounds screened with an overall score of +15.86. ICM likely docked this compound outside of the active site of DDl, but this contradictory information raises the question of both screening softwares' capability to correctly predict ligand docking poses as well as this compound's actual binding abilities.

GOLD Rank	ICM Rank	Library	GOLD Score	ICM Score	S(PLP)	S(hbond)	Ligand Name	Ligand Efficiency	MW	Log P	Hbond Donor	Hbond Acceptor	Lipinski?
1	23	ZINC	92.22	15.86	-76.95	6.00	ZINC_1	0.209	441	3.00	~	~	~
2	18	ZINC	92.18	-2.66	-84.44	2.87	ZINC_2	0.202	457	4.47	2	4	Yes
3	~	ZINC	90.85	~	-81.27	3.98	ZINC_3	0.188	484	2.54	1	6	Yes
4	~	ZINC	90.27	~	-79.85	2.70	ZINC_4	0.189	477	3.05	0	5	Yes
5	~	ZINC	90.2	~	-86.26	1.70	ZINC_5	0.191	472	2.90	0	6	Yes
6	14	HF9	87.8	-9.41	-84.01	1.84	HF9_1	0.185	474	~	~	~	~
7	9	HF9	86.39	-15.66	-82.35	1.92	HF9_2	0.195	443	~	~	~	~
8	21	HF9	85.2	5.27	-81.53	1.72	HF9_3	~	~	~	~	~	~
9	11	HF9	84.79	-11.94	-83.04	1.97	HF9_4	0.143	591	~	~	~	~
10	22	HF9	84.64	5.53	-81.18	2.57	HF9_5	0.224	378	~	~	~	~
11	15	IHC	80.55	-8.32	-71.62	2.85	IHC_1	0.172	467	3.20	2	5	Yes
12	8	CB306	76.88	-18.26	-66.02	2.90	CB306_1	0.202	380	0.65	5	4	Yes
13	20	IHC	73.47	0.82	-70.36	2.00	IHC_2	0.194	379	2.09	3	3	Yes
14	12	IHC	73.25	-11.82	-68.05	2.42	IHC_3	0.194	377	2.69	1	5	Yes
15	16	IHC	73.25	-6.53	-68.94	1.98	IHC_4	0.219	335	2.42	3	2	Yes
16	13	IHC	72.66	-11.27	-68.99	1.71	IHC_5	0.193	377	2.69	1	5	Yes
17	17	CB306	72.05	-4.45	-64.59	1.77	CB306_2	0.188	383	2.20	0	6	Yes
18	19	CB306	71.37	-0.33	-66.33	1.00	CB306_3	0.183	391	1.72	1	5	Yes
19	~	CB306	70.56	~	-59.52	3.08	CB306_4	0.239	295	1.68	4	4	Yes
20	10	CB306	69.32	-14.71	-59.67	3.84	CB306_5	0.203	342	2.70	2	7	Yes
21	1	ZINC	67.78	-34.72	-54.95	4.46	ZINC_6	0.251	270	2.05	3	3	Yes
22	7	HF9	65.64	-25.22	-56.17	2.45	HF9_9	0.159	413	~	~	~	~
23	2	ZINC	63.72	-32.21	-51.61	4.27	ZINC_7	0.215	297	2.91	2	5	Yes
24	3	CB306	61.5	-30.20	-51.13	3.51	CB306_6	0.187	329	1.79	0	5	Yes
25	4	HF9	50.91	-29.42	-45.61	1.88	HF9_6	0.280	182	~	~	~	~
26	5	HF9	49.83	-27.64	-46.63	1.74	HF9_7	0.191	261	~	~	~	~
27	6	HF9	48.67	-26.99	-42.17	0.32	HF9_8	0.158	309	~	~	~	~

Table 2: Top scoring experimental ligands docked with GOLD and ICM.

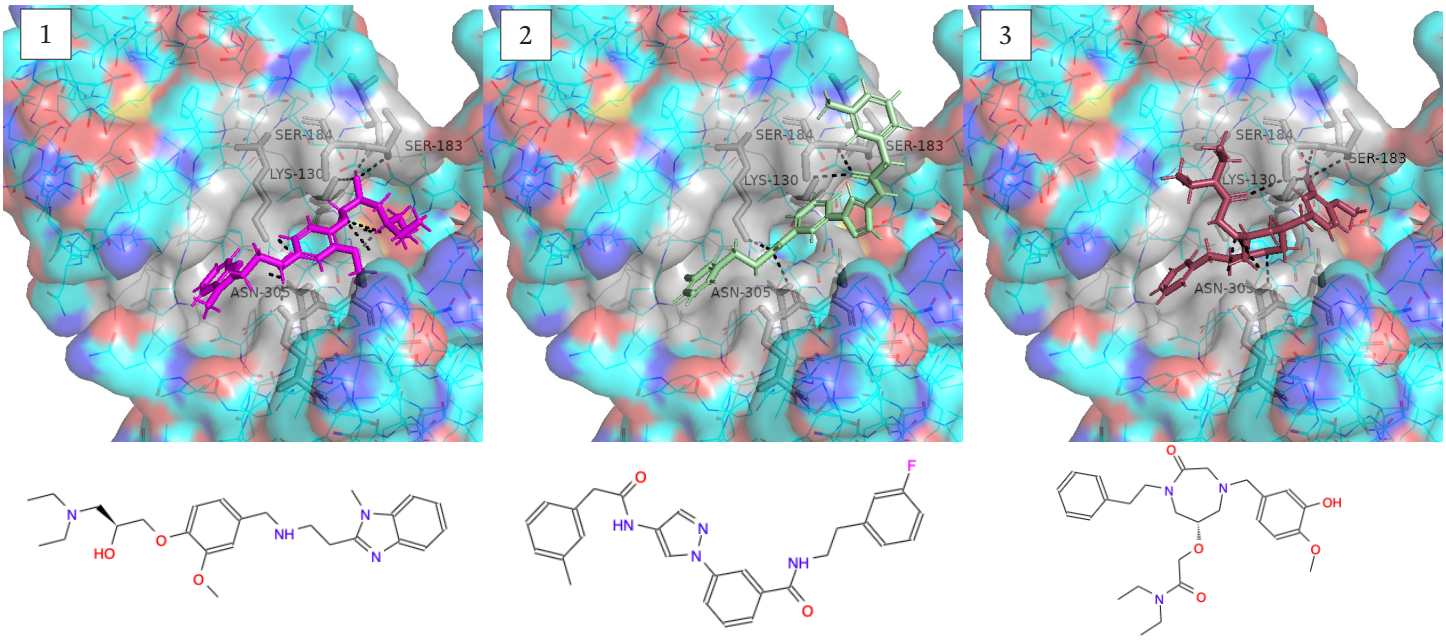


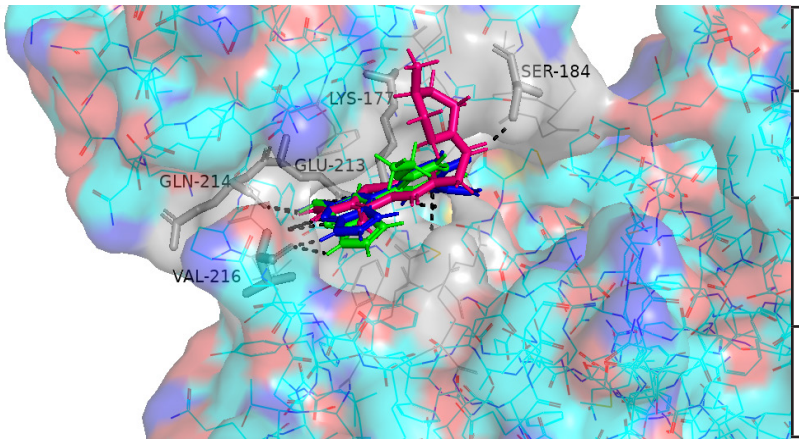
Figure 6: Top three ranked GOLD docked experimental ligands. The DDl protein is shown as lines with a semi-transparent surface and carbons colored light blue. The docking site is colored grey and important amino acids for docking are labeled and shown as grey sticks. The docked ligands are shown as sticks with polar contacts indicated by black dashed lines. Box 1 shows compound ZINC_1, box 2 compound ZINC_2, box 3 compound ZINC_3[10]. The 2D structure of the compounds is shown below the 3D pymol image[10].

Some ligands, like CB306_1, received moderate scores from both GOLD and ICM docking. CB306_1 had the best ICM score (-18.26) out of the top-ranked GOLD compounds, yet was not low enough to pass the -25 ICM score threshold used for screening the CB306 library. This ligand received a GOLD score of 76.88 and the score improved significantly when normalized, with a ligand efficiency score of 0.202. The disparities in overall GOLD and ICM scores may be due to the molecules' somewhat small size, and the compound may succeed in a wet lab better than virtual screening anticipated.

ICM Results

ICM provided two ligands (ZINC_6 and ZINC_7) out of 124,271 from the ZINC library with scores below the -32 threshold, however, there likely would have been more results if a larger threshold was used[10]. Neither of these two ligands were included in the top 5 GOLD ranked ZINC ligands, however, both received overall scores in the top 35% of all compounds screened with GOLD. The CB306 library had one compound, CB306_6, with an ICM score below the -25 threshold and received a score of -30.20. This compound obtained a GOLD score of 61.50, which is above the average ligand's overall GOLD score but does not indicate a binding potential as strong as the ICM score alone would predict. HF9 had the most compounds scoring below the -25 threshold with four ligands scoring between -25.22 and -29.42. The IHC library had no compounds with ICM scores below the indicated threshold.

The top ICM ranked compound, ZINC_6, received a very good ICM score of -34.72. This compound was in the top 25% of scores for all compounds screened with GOLD with an overall binding score of 67.78. This provides strong evidence that this compound would likely bind well in vitro environments.
The top three ICM ranked compounds (ZINC_6, ZINC_7, and CB306_6) had similar docking poses and polar contacts. These top- ranked compounds all interacted with Val-216, Gln-214, and Glu-213. Additionally, CB306_6 and all four of the top ICM ranked HF9 ligands had polar contacts with Lys-177. While these amino acids may be important for binding or protein folding, these are not the same amino acids that GOLD predicted as polar contacts. The virtual screening softwares' scores and poses are estimates and one software's algorithm may be more accurate than the other. Likewise, the predicted docking location may not be exactly where a ligand



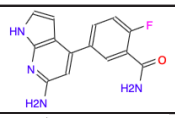
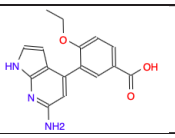
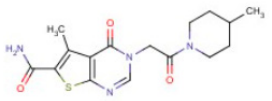
Ligand Name (color)	2D Structure
ZINC_6 (green)	
ZINC_7 (blue)	
CB306_6 (pink)	

Figure 7: Top-scoring ICM docked ligands. The DDl protein is shown as lines with a semi-transparent surface and carbons colored light blue. The docking site is colored grey and important amino acids for docking are labeled and shown as grey sticks. The docked ligands are shown as sticks with polar contacts indicated by black dashed lines. The ZINC_6 ligand is colored green, ZINC_7 colored dark blue, and CB306_6 colored pink. The 2D structures of each ligand is shown to the right of the PyMol image[10,17].

would bind in vitro and could explain the discrepancy.

GOLD vs. ICM Results

Overall, GOLD and ICM did not result in similar scores as expected. None of the top-ranked compounds determined by GOLD received scores below the ICM score threshold of -25, or -32 for the ZINC library. A majority of these ligands did not have ICM scores below -10, and several of the top 20 GOLD ranked compounds even received positive (unfavorable) ICM scores. Likewise, none of the top-ranked ICM compounds received GOLD scores above 70. These results were unexpected as we hoped to obtain favorable scores from both virtual screening softwares to provide strong evidence that these compounds would bind well when docked against DDl in the lab. The score differences between the two virtual screening softwares was due to the variations in docking algorithms and one program likely has more accurate binding predictions than the other. However, the accuracy of binding predictions cannot be measured until the top scoring ligands from both GOLD and ICM are tested with inhibition assays.

CONCLUSION AND FUTURE WORK

Virtual screening cannot definitively prove whether or not a compound will have strong inhibitory abilities when tested in a wet lab, but does provide a good starting point for which compounds would likely succeed *in vitro* conditions. The next steps of this project include completing molecular dynamics simulations to provide more intensive computational predictions of binding affinities to better refine the top compounds list. Before the top compounds can be experimentally tested, the coding DNA sequence of DDl will be cloned, the DDl protein will be expressed, and the protein's enzymatic activity will be measured through spectrophotometric assays. The top-scoring ligands determined by virtual screening, such as ZINC_1 or ZINC_6, will be tested for their inhibitory abilities against D-alanine: D-alanine Ligase with these biochemical assays.

The wet lab results may indicate which screening software provides more accurate results and guide future work to develop a compound into a lead compound towards therapeutic purposes. Additionally, testing compounds like CB306_1 could provide useful insight about whether or moderate scores from both softwares provides evidence of relatively strong binding potential compared to a favorable score from one software and an unfavorable score from the other screening software. This comparison would help validate the software programs to allow further use in drug discovery efforts with this enzyme in different types of bacteria.

ACKNOWLEDGEMENTS

A special thanks to Dr. Josh Beckham for his continued support and mentorship while completing this research. This work would not have been possible without resources provided by the Virtual Drug Screening (VDS) Stream of the Freshman Research Initiative. Thank you to Dr. Walter Fast, the principal investigator of the VDS laboratory. I would also like to thank the Office for Undergraduate Research for providing funding for this work through the Undergraduate Research Fellowship.

SUPPLEMENTAL IMAGES

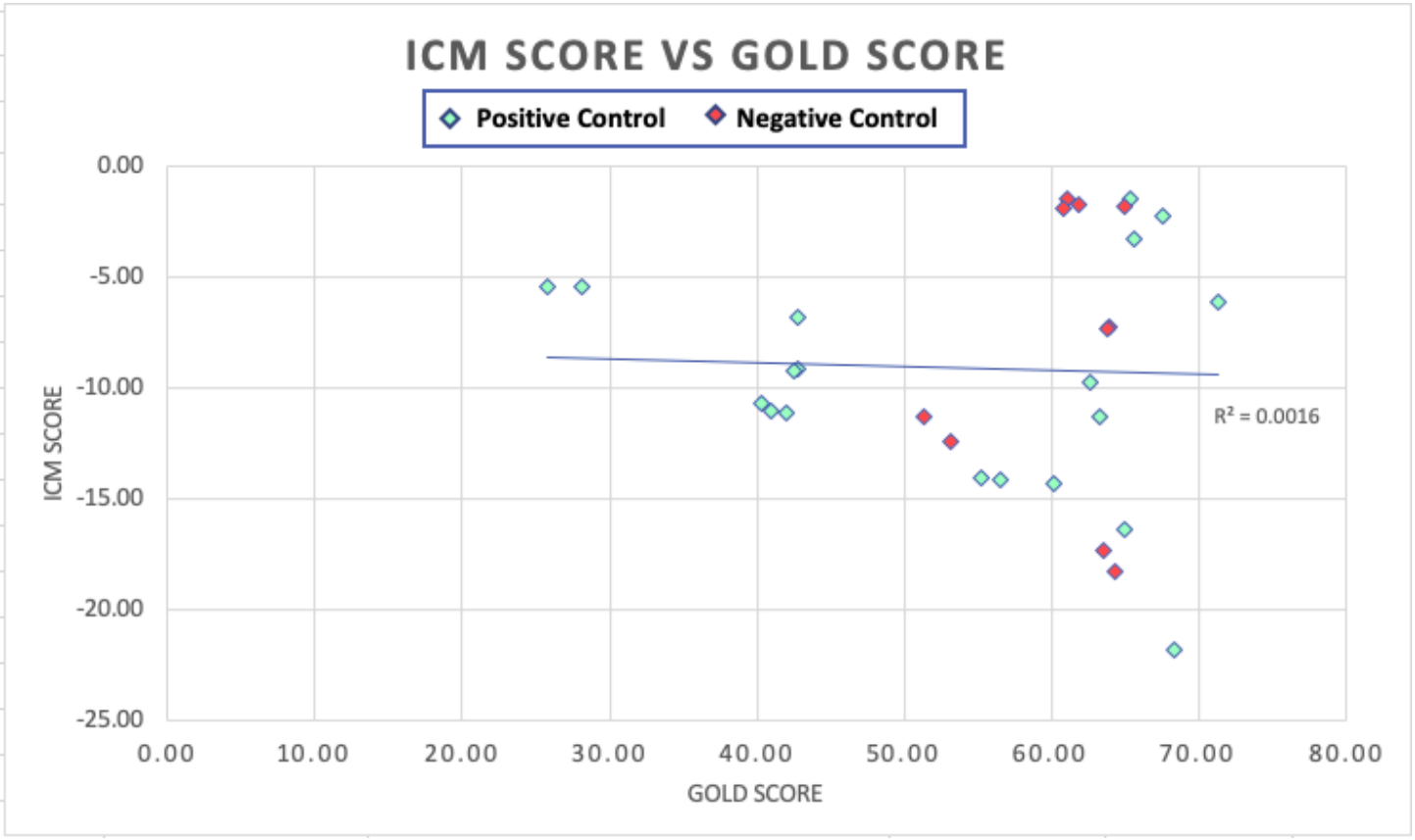


Figure S1: Plot of ICM generated control ligand binding scores versus GOLD generated control ligand binding score. The positive controls are shown as green diamonds and the negative controls are shown as red diamonds. The blue line represents the trend line with an R2 value of 0.0016.

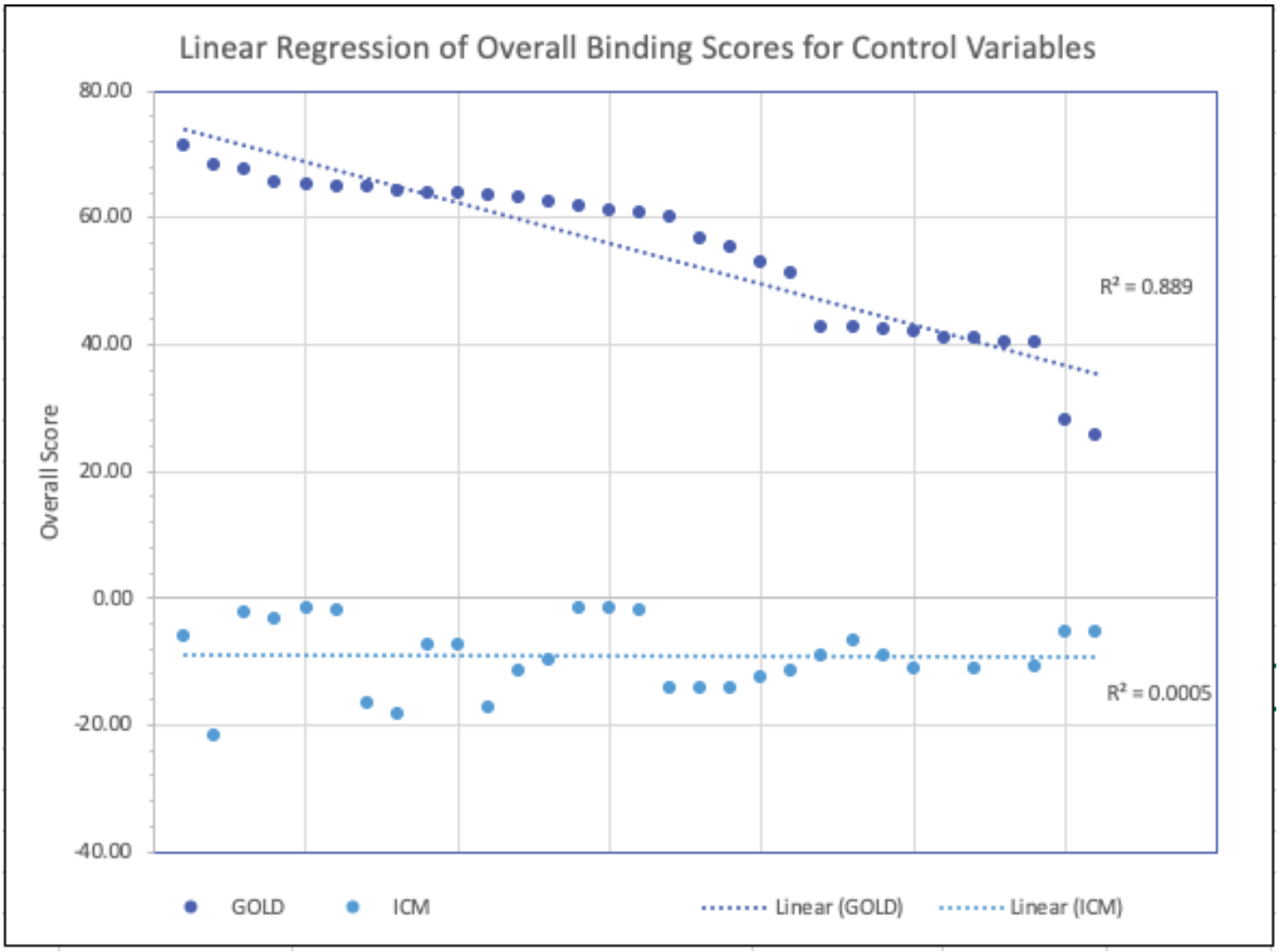


Figure S2: Plot of overall binding scores for control variables. GOLD results are shown as dark blue dots while ICM results are shown as light blue dots. The GOLD generated results have a stronger correlation (R^2=0.8890) compared to ICM results (R^2=0.0005).

REFERENCES

1. David, M. Z; Daum, R. S. Community-associated methicillin-resistant *Staphylococcus aureus*: epidemiology and clinical consequences of an emerging epidemic. *Clin Microbiol Rev*. 2010. 23, (3), 616-687.

2. Patel, S.; Preuss, C. V.; Bernice, F. Vancomycin. *StatPearls*. 2020.

3. Lowy, F. D. *Staphylococcus aureus* Infections. *The New England Journal of Medicine.* 1998. 339, 520-532.

4. Liu, S.; Chang, J. S.; Herberg, J. T.; Horng, M.; Tomich, P. K.; Lin, A. H.; Marotti, K. R. Allosteric Inhibition of *Staphylococcus aureus* D-alanine:D-alanine Ligase Revealed by Crystallographic Studies. *Proceedings of the National Academy of Sciences.* 2006. 103, (41), 15178-15183.

5. Center for Disease Control and Prevention (CDC). Staph Infections can Kill. 2019.

6. Chambers, H. F.; Deleo F.R. Waves of resistance: Staphylococcus aureus in the antibiotic era. *Nat Rev Microbiol*. 2009;7(9):629-641.

7. Center for Disease Control and Prevention (CDC). *Staphylococcus aureus* in Healthcare Settings. 2011.

8. Centers for Disease Control and Prevention (CDC). Outbreaks of community-associated methicillin-resistant Staphylococcus aureus skin infections--Los Angeles County, California, 2002-2003. MMWR Morb. Mortal. Wkly. 52(5):88.

9. Center for Disease Control and Prevention (CDC). *Staphylococcus aureus* in Healthcare Settings. 2011.

10. Sterling T.; Irwin, J.J. ZINC-15 Ligand Discovery for Everyone. *Journal of Chemical Information and Modeling.* 2015, 55, (11), 2324-2337.

11. Taylor, T. A.; Unakal, C. G. *Staphylococcus aureus. StatPearls.* 2019.

12. Jones G., Willett P. and Glen R. C., Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation, *J. Mol. Biol*., 1995, 245, 43-53.

13. Neves, M.; Totrov, M.; Abagyan, R. Docking and scoring with ICM: the benchmarking results and strategies for improvement. *J. Comput.* Aided. Mol. Des. 2012, 26, (6), 675-686.

14. Meziane-Cherif D.; Saul F.A.; Moubareck C.; Weber P.; Haouz A.; Courvalin P.; Périchon B. Molecular basis of vancomycin dependence in VanA-type Staphylococcus aureus VRSA-9. *J Bacteriol*. 2010, 192, (20), 5465-5471.

15. PubChem [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; 2004-. PubChem Compound Summary for CID 33113, Phosphoaminophosphonic acid-adenylate ester; [cited 2021 Apr. 8].

16. Vincent B. Chen, W. Bryan Arendall III, Jeffrey J. Headd, Daniel A. Keedy, Robert M. Immormino, Gary J. Kapral, Laura W. Murray, Jane S. Richardson and David C. Richardson (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallographica D66: 12-21.

17. Chembridge. Hit2Lead – Chembridge Online Chemical Store. https://www.hit2lead.com (accessed April 15, 2021).

18. ThermoFisher Scientific. Maybridge Hitfinder Screening Library. https://www.thermofisher.com/order/catalog/product/HF13000?us&en#/HF13000?us&en (accessed April 15, 2021).

19. Boudrioua, A; Li, Y.; Hartke, A.; Giraud, C. Opposite effect of vancomycin and D-Cycloserine combination in both vancomycin resistant *Staphylococcus aureus* and enterococci. *FEMS Microbiology* Letters. 2020, 367, (8).

20. Batson, S.; de Chiara, C.; Majce, V. *et al.* Inhibition of D-Ala:D-Ala ligase through a phosphorylated form of the antibiotic D-cycloserine. *Nat Commun* 8, 1939 (2017).

21. Prosser, G.A.; de Carvalho, L.P. Reinterpreting the mechanism of inhibition of Mycobacterium tuberculosis D-alanine:D-alanine ligase by D-cycloserine *Biochemistry*. 2015. 52, (40), 7145-7149.

Motivational Interviewing, Self-efficacy, and Depressive Symptoms

Madeline Harrington
Special Honors in the Department of Psychology, The University of Texas at Austin
Texas Undergraduate Research Journal

ABSTRACT

Motivational interviewing (MI), a therapeutic method, seeks to draw out intrinsic motivation in order to change a particular behavior. To do this, one of MI's main goals is to support self-efficacy. Elevated depressive symptoms are linked to decreased self-efficacy, so if MI is successful in increasing self-efficacy, there could be hopeful implications for decreasing depressive symptoms. Data for the present study comes from a parent study evaluating an MI-based intervention that reduced the risk of alcohol-exposed and/or tobacco-exposed pregnancies. Participants (n = 261) were randomly assigned to two experimental groups (MI and control), and self-efficacy and depressive symptoms were measured at baseline (pre-intervention) and a 3-month and 9-month follow-up. Depressive symptoms decreased significantly over time for both groups. Self-efficacy was significantly negatively correlated with depressive symptoms, and MI predicted a significant increase in self-efficacy regarding alcohol use at both follow-ups. A significant indirect effect was found of MI on depressive symptoms through self-efficacy regarding alcohol use at both follow-ups, partially confirming the hypothesized mediational model. Additional analyses examined symptoms of anxiety, symptoms of somatization, and general symptom severity, revealing similar indirect effects of MI. The present study implies MI's utility for improving various symptoms both for those with behavior-based problems and those seeking treatment for other mental health problems.

INTRODUCTION

Depression is an extremely common mental disorder, with a lifetime prevalence of Major Depressive Disorder (MDD) of 20% in U.S. adults (Hasin, Sarvet, Meyers, Ruan, Stohl, & Grant, 2018), and the number of people with subclinical, yet still worrisome and hindering depressive symptoms is much larger still. Effective treatments for depressive symptoms (such as Cognitive-Behavioral Therapy, or CBT) are common (Honyashiki, Furukawa, Noma, Tanaka, Chen, Ichikawa, Ono, Churchill, Hunot, & Caldwell, 2014), but their effects do not always last or protect against recurrence (Hollon, Jarrett, Nierenberg, Thase, Trivedi, & Rush, 2005). To improve outcomes, especially in the long-term, supplementing these treatments with another treatment may be a fitting solution. Among supplemental treatments that have shown promise is motivational interviewing.

Motivational interviewing (MI) has been primarily used to motivate people to change specific behaviors and has been well-supported for this purpose, but its effects on cognition are not quite so clear. This may mean that this therapeutic strategy may be underutilized for clients with primarily depressive symptomatology. The purpose of the current study is to examine: (a) the effects of motivational interviewing on depressive symptoms, and (b) the role

of self-efficacy in mediating this effect. A graphic representation of this design can be seen in Figure 1. What exactly is motivational interviewing (MI), and how does it work? Developed from Carl Rogers's person-centered perspective (Miller & Moyers, 2017), MI is a method of communication that seeks to draw out intrinsic motivation from people and direct it toward changing a particular behavior (Miller & Rollnick, 2002). Many people partake in problematic behaviors and are even aware of their harmful effects, but have difficulty changing these behaviors. Often, this difficulty arises from feelings of ambivalence toward the behavior; the person is both attracted to and disapproving of the behavior (Miller & Rollnick, 2002). Intuitively, it may seem like the best way to motivate change would be to provide a convincing argument for change. However, in a directive counseling style, clients often respond to such arguments by presenting a counterargument for why *not* to change (Westra & Aviram, 2013). Therefore, although it may seem productive to provide someone with reasons to change, this is not always an effective way to actually produce change.

Alternately, MI seeks to create change by drawing out motivation and confidence from within the client. A central tenet and main goal of MI is supporting self-efficacy in the client. In other words, it is important that the client themself believes that they can change. MI has been shown to successfully increase self-efficacy (Berman, Forsberg, Durbeej, Källmén, & Hermansson, 2010), and Huang and colleagues highlighted the enhancement of self-efficacy as *the* crucial benefit of MI for their study on lung cancer patients (Huang, Yang, Zhang, Han, Zhang, & Ye, 2018). An exploratory pilot study conducted by Khattra, Angus, Westra, Macaulay, Moertl, and Constantino (2017) done with two clients who recovered from Generalized Anxiety Disorder (GAD) suggests that self-efficacy could have important implications for clients' outcomes post-therapy and their maintenance of gains made in therapy. For example, one client (given the pseudonym "Deb"), who received CBT treatment, mainly attributed her recovery to the acquisition of CBT skills as taught to her by her therapist. On the other hand, the client ("Martha") who received CBT with MI seemed to attribute her recovery to her own hard work, improved self-understanding, and capability in finding answers to her problems. This language implies positive perceptions of her own self-efficacy as being a producer of change. Additionally, Martha maintained her status as recovered from GAD at a one-year follow-up, while Deb had relapsed at the one-year mark. These findings, though exciting, cannot be generalizable due to the small sample size (n=2) of the study and differences between the two participants and the two therapists, but they still offer a promising direction for future research regarding self-efficacy as a mechanism of MI.

Motivational interviewing has been shown to be useful in a wide variety of contexts and with various populations (Frost, Campbell, Maxwell, O'Carroll, Dombrowski, Williams, Cheyne, Coles, & Pollock, 2018; Burke, Arkowitz, & Menchola, 2003). MI was originally created as a treatment for alcohol dependence (Miller & Rollnick, 2002) and has been particularly researched in the realm of substance abuse and related disorders and health-related behaviors (Arkkukangas et. al., 2018; D'Amico, Houck, Tucker, Ewing & Pedersen, 2017; Pfund, Whelan, Peter, & Meyers, 2020). Consistently, MI seems to be effective at reducing addiction-related behaviors, such as alcohol use and gambling (Borsari et. al., 2019, Magill et. al., 2019, Pfund et. al., 2020). In the area of health behaviors, previous research has also focused on behaviors such as adherence to exercise (Arkkukangas et. al., 2018). In the areas of internalizing disorders such as depression and anxiety, however, research is less extensive, although there has been progress regarding this question in the past two decades (Keeley, Brody, Engel, Burke, Nordstrom, Moralez, Dickinson, & Emsermann, 2016; Muir, Constantino, Coyne, Westra, & Antony, 2019; Westra, Arkowitz, & Dozois, 2009; Arkowitz, Westra, Miller, & Rollnick, 2007). Though this research has shown that MI does seem to be helpful in treating such prevalent disorders as depression and anxiety even when they are not comorbid with behavioral disorders (such as addictions-related disorders), the mechanisms of this relationship and implications for long-term outcomes are not yet fully understood.

Moreover, existing research on MI as a treatment for internalizing disorders such as anxiety and depression often measures self-efficacy as a part of

this relationship. In a study done with lung cancer patients, a self-efficacy enhancement treatment based on MI helped reduce anxiety and depression, in addition to improving self-efficacy (Huang et. al., 2018). In another study, Worley, Tate, Granholm, and Brown (2014) found that in veterans with neurocognitive impairments diagnosed with both alcohol dependence and Major Depressive Disorder (MDD), for participants with severe impairment, depressive symptoms moderated the effect of 12-step program affiliation on the percentage of days drinking such that more severe depression and higher 12-step affiliation more strongly predicted lower future drinking. Similarly, depressive symptoms interacted with self-efficacy for severely impaired participants to predict future drinking, with more severe depressive symptoms increasing the effect of greater self-efficacy on lowering future drinking. Though these moderating effects of depressive symptoms were only found in participants with severe neurocognitive impairment, this study may offer evidence that supports the effects of MI on depressive symptoms through self-efficacy in the present study.

PRESENT STUDY

The present study will examine how MI relates to self-efficacy and, in turn, how this leads to changes in depressive symptoms. MI, in conjunction with CBT, is not often a primary tool used for clients with primarily depressive symptomatology, but research where MI has been used for these and similar issues, as well as MI's effects on self-efficacy, suggests that it may be useful for problematic depressive symptoms both when they do and do not occur in conjunction with a primarily behavioral disorder. For those seeking treatment for a primarily behavior-based problem, MI could also incur broader benefits in surplus of changing the target behavior. In addition, if MI indeed fosters a narrative in which the client feels that they have been able to successfully produce a positive change by their own agency, there may be positive implications for attitudes when encountering problems in the future and the durability of change. For the present study, motivational interviewing was predicted to decrease depressive symptoms. In addition, perceptions of self-efficacy were predicted to mediate the effect of motivational interviewing on depressive symptoms.

METHOD

Sample Selection and Characteristics

The present study is part of an overall parent study conducted from 2011-2013 that examined the effectiveness of a motivational interviewing-based intervention called Choices Plus, whose goal was to reduce the risk of having an alcohol-exposed pregnancy (AEP) or a tobacco-exposed pregnancy (TEP) (Velasquez, Sternberg, Floyd, Parrish, Kowalchuk, Stephens, ... Mullen, 2017). Inclusion criteria for the parent study included the following: must be a fertile woman, 18-44 years of age, not pregnant or planning to become pregnant, must have had vaginal intercourse with ineffective contraception in the past three months, and must have drank at risky levels in the past three months. 45% of the participants in this study (n= 118) were also tobacco users, although tobacco use was not a criterion for eligibility in this study.

Women were recruited for the Choices Plus program from primary care settings in Harris County, Texas, and by posters placed in clinic and hospital waiting rooms. Data were collected between April 2011 and October 2013, and participants were measured at baseline, at in-person follow-ups after three months and nine months, and in a telephone follow-up after six months. Women received $75 for original baseline measurement and $30 for the 3-month follow-up, and women in the MI-based intervention group also received $30 for attending the second intervention session. The study protocol was approved by IRBs at the University of Texas at Austin, Baylor College of Medicine, and the Harris Health System. For additional information on the sample recruitment and data collection procedure, see Velasquez et al. (2017).

DESIGN

The present study examined self-efficacy and depressive symptoms in a mixed design. Motivational interviewing was manipulated between subjects at two levels (motivational interviewing and control), and self-efficacy and depressive symptoms were measured within subjects at three levels (at baseline pre-intervention, 3-month follow up, and 9-month

follow-up). Participants were randomly assigned to the two experimental groups (motivational interviewing and control). The experimental group received an MI-based intervention called Choices Plus conducted by a Behavioral Health Specialist, and the control group received a brief advice session with a Behavioral Health Specialist. The presence or absence of the MI intervention was examined as the main predictor, and depressive symptoms were examined as a continuous variable as the main outcome. Self-efficacy was assessed regarding each of two or three behaviors: alcohol use, contraception use, and tobacco use (only the participants who were identified as smokers were assessed for self-efficacy regarding tobacco use). Self-efficacy was examined as a continuous variable as the proposed mediator.

EXPERIMENTAL MANIPULATION

Motivational Interviewing

Motivational interviewing was manipulated by the MI-based Choices Plus intervention as outlined in a manual developed for the parent study. The intervention was delivered by trained masters' level Behavioral Health Specialists (BHSs). All BHSs were proficient in motivational interviewing, and the training program for Choices Plus also included a foundation of MI theory and practice. The intervention consisted of two 40-minute sessions. The first session was conducted after the baseline interview, and the second intervention was conducted approximately two weeks later (before the 3-month follow-up).

Control

Participants in the control group received a brief advice session about their alcohol and, if applicable, tobacco use delivered by BHSs. They were also given a "Healthy Lifestyles" brochure, a referral brochure to local community services, and referrals to a local health services site that provides contraceptives, tobacco, alcohol, and other drug services. For additional information about the manipulation and control conditions and procedures, see Velasquez et al. (2017).

MEASURES

Self-efficacy

Self-efficacy was assessed by a 43-item questionnaire based on the Situational Confidence Questionnaire (Miller, Ross, Emmerson, & Todt, 1989) that measured participants' temptation and confidence regarding specified behaviors in certain situations. The confidence dimension of the questionnaire represents participants' self-efficacy (DiClemente, Prochaska, & Gibertini, 1985; Breslin, Sobell, Sobell, & Agrawal, 2000; Velasquez, et., al., 2017). Of the 43 items, 21 measured confidence about their behaviors (7 questions related to alcohol use, 5 questions related to contraceptive use, and 9 questions related to tobacco use), and 22 measured the temptation of the behaviors (8 questions related to alcohol use, 5 questions related to contraceptive use, and 9 questions related to tobacco use). Each item asked participants to rate their degree of confidence regarding their ability to resist a behavior in a given situation or their temptation toward a behavior in a given situation on a scale from 1-5. Items for self-efficacy (or confidence) regarding each behavior were averaged to yield the scores we used to represent self-efficacy so that a score of 1 represents no self-efficacy and a score of 5 represents maximal self-efficacy. This questionnaire was administered at all three time points (baseline pre-intervention, 3-month follow-up, and 9-month follow-up).

Depressive Symptoms

The Brief Symptom Inventory-18 (BSI-18) was included to broadly measure psychological symptoms, including depression (Derogatis, 1993). The BSI-18 includes 18 items relating to three different symptom categories, Somatization (SOM), Depression (DEP), and Anxiety (ANX). Participants respond to these items (e.g., "your feelings being easily hurt") on a 5-point scale (0-4) from low to high intensity. The BSI-18 also produces an index that indicates the severity of symptoms experienced called the Global Severity Index (GSI). For the three subscales and the GSI, higher scores reflect higher distress. Items from each subscale of the BSI-18 were averaged to yield the number we used to represent symptomatology, so that a score 0 represents no symptomology and

a score of 4 represents the most severe symptomatology. This tool has good reliability in the range of 0.7. The BSI-18 was administered at all three time points.

Manipulation Check

In the parent study, Choices Plus sessions were audio recorded and reviewed by supervisors to ensure quality, and supervisors also rated the BHSs' skillfulness of MI with the Motivational Interviewing Treatment Integrity Scale 3.1.133 (MITI 3) (Moyers, Martin, Manuel, Miller, & Ernst, 2010). This tool has been shown to have a good-to-excellent reliability (Pierson, Hayes, Gifford, Roget, Padilla, Bissett, Berry, Kohlenberg, Rhode, & Fisher, 2007).

STATISTICAL PROCEDURE

The data were analyzed using repeated measures ANOVAs to look at changes in self-efficacy, depressive symptoms, symptoms of anxiety, symptoms of somatization, and general symptom severity over time across and between conditions. Pearson correlational analyses assessed the relationships between self-efficacy, depressive symptoms, symptoms of anxiety, symptoms of somatization, and general symptom severity. One-way ANOVAs assessed the effect of MI on self-efficacy, depressive symptoms, symptoms of anxiety, symptoms of somatization, and general symptom severity at each static time point for the 3-month and the 9-month follow-ups. We used linear regression to assess the effect of condition on depressive symptoms, with the presence or absence of the Choices Plus MI-based intervention as a predictor, self-efficacy regarding each of the three assessed behaviors as a mediator, and depressive symptoms as the main outcome. Significance of mediation was tested with the PROCESS macro (Hayes, 2018).

RESULTS

The present study analyzed data from a total of 261 participants in the parent study. Participants from the parent study had a mean age of 31 years and were largely Hispanic (47.1%) or non-Hispanic black (41.8%). A majority of participants had household incomes of <$20,000 (70.7%) at the time of measurement, and 40.6 % were married or lived with a partner. Since a relatively small number of participants had incomplete data for one or more measures, analyses varied slightly in the number of data points included from about n = 220 to n = 261.

MAIN EFFECTS

The present study hypothesized that motivational interviewing would decrease depressive symptoms. In addition, perceptions of self-efficacy were predicted to mediate the effect of motivational interviewing on depressive symptoms. Self-efficacy for three different behaviors (alcohol use, contraception use, and tobacco use) and depressive symptoms were measured at baseline (pre-intervention), a 3-month follow-up, and a 9-month follow-up for participants in both the motivational interviewing treatment group and the control brief advice group.

Depressive Symptoms

Repeated measures ANOVA showed a significant linear reduction in depressive symptoms over time for both groups ($F = 36.29$, $df = 1$, $p < .001$), but this effect did not significantly differ between conditions ($F = .29$, $df = 1$, $p = .59$). A test of linear regression controlling for initial reported depressive symptoms as measured at baseline showed that level of depressive symptoms did not moderate the effect of condition on depressive symptoms at the 3-month follow-up ($\beta = -.01$, $p = .87$) or the 9-month follow-up ($\beta = .04$, $p = .62$).

Self-efficacy

A general linear model using repeated measures ANOVA revealed an overall significant between-subjects increase for self-efficacy regarding alcohol use ($F = 8.30$, $df = 1$, $p < .01$). The model also showed an overall significant linear increase in self-efficacy regarding alcohol use over time within subjects across both conditions ($F = 10.23$, $df = 1$, $p < .01$). When we tested the general linear model using repeated measures ANOVA for self-efficacy regarding contraception use, it revealed a significant linear increase over time for both conditions within subjects ($F = 5.61$, $df = 1$, $p = .02$), but did not show a significant effect between subjects ($F = .83$, $df = 1$, $p = .37$). When

the general linear model was tested with self-efficacy regarding tobacco use, the repeated measures ANOVA revealed a marginal linear increase over time for both groups ($F = 3.61$, $df = 1$, $p = .06$), but did not find an overall difference in self-efficacy regarding tobacco use over time between subjects ($F = .27$, $df = 1$, $p = .61$).

Self-efficacy x Depressive Symptoms

Pearson correlational analyses of self-efficacy regarding each of the three behaviors with depressive symptoms revealed weak negative correlation between the two constructs, with higher self-efficacy predicting lower depressive symptoms. These correlations are presented in Table 1, and they were significant at all three time points for all three behaviors, confirming the hypothesized relationship between self-efficacy and depressive symptoms.

Mediational Model

One-way ANOVA analyses showed that self-efficacy regarding alcohol use was significantly higher for the motivational interviewing group at both the 3-month follow-up ($F = 1.57$, $df = 1$, $p = .03$) and the 9-month follow-up ($F = 2.52$, $df = 1$, $p < .001$). Based on these analyses, we only tested the hypothesized mediational model for self-efficacy regarding alcohol use. Consistent with the hypothesized model (see Figure 1), motivational interviewing predicted lower depressive symptoms indirectly through self-efficacy regarding alcohol use at both follow-ups. We confirmed significant mediation with percentile bootstrap confidence intervals (CIs) of the indirect effect using the PROCESS macro at the 3-month (indirect effect = -.10; 95% CI [-.20, -.03]) and the 9-month follow-ups (indirect effect = -.08; 95% CI [-.15, -.008]; Hayes, 2018).

ADDITIONAL ANALYSES

Symptoms of Anxiety and Somatization

The BSI-18 measures symptoms of anxiety and somatization as well as depression (Derogatis, 1993), so the present study also looked at the relationship between motivational interviewing, self-efficacy, and these symptom measures. Table 2 presents the means for the additional symptom measures by condition. A general linear model using repeated measures ANOVA found that anxiety symptoms reduced significantly over time for both conditions ($F = 22.27$, $df = 1$, $p < .001$), but there was not a significant difference between conditions ($F = .23$, $df = 1$, $p = .63$). When the general linear model using repeated measures ANOVA was tested with symptoms of somatization, it revealed a significant linear decrease in symptoms of somatization over time for both conditions ($F = 5.07$, $df = 1$, $p = .03$) and a significant quadratic decrease in symptoms of somatization over time for both conditions ($F = 4.23$, $df = 1$, $p = .04$). The general linear model also showed a weak marginal decrease in symptoms of somatization over time between subjects ($F = 2.29$, $df = 1$, $p = .13$).

One-way ANOVAs found that self-efficacy regarding alcohol use predicted a significant decrease in symptoms of anxiety at the 3-month follow-up ($F = 1.61$, $df = 1$, $p = .03$) and a weak marginal decrease in symptoms of anxiety at the 9-month follow-up ($F = 1.36$, $df = 1$, $p = .11$). One-way ANOVA also showed that self-efficacy regarding tobacco use marginally predicted a decrease in symptoms of anxiety at the 3-month follow-up ($F = 1.45$, $df = 1$, $p = .10$). When tested with symptoms of somatization, one-way ANOVA showed that self-efficacy regarding alcohol use significantly predicted a decrease in symptoms of somatization at the 9-month follow-up ($F = 1.64$, $df = 1$, $p = .03$) and that self-efficacy regarding contraception use marginally predicted a decrease in symptoms of somatization at the 9-month follow-up ($F = 1.44$, $df = 1$, $p = .10$).

Pearson correlational analyses showed a weak significant negative relationship between self-efficacy and symptoms of anxiety for all three behaviors at baseline and the 3-month follow-up and for self-efficacy regarding alcohol use and tobacco use at the 9-month follow-up. A weak significant negative correlation was found between symptoms of somatization and self-efficacy regarding alcohol use for all three time points, between symptoms of somatization and self-efficacy regarding contraception use at baseline, and between symptoms of somatization and self-efficacy regarding tobacco use at the 3-month and 9-month follow-ups. The coefficients

of these correlations are shown in Table 3.
Based on the above analyses, the present study tested the indirect effect of motivational interviewing on symptoms of anxiety through self-efficacy regarding alcohol use for both follow-ups. Consistent with this modified mediational model, motivational interviewing predicted a reduction in symptoms of anxiety indirectly through self-efficacy regarding alcohol use at the 3-month follow-up, but the indirect effect was not significant at the 9-month follow-up. Using the PROCESS macro, we confirmed significant mediation with percentile bootstrap confidence intervals (CIs) of the indirect effect at the 3-month follow-up (indirect effect = -.08; 95% CI [-.17, -.02]) and non-significant marginal mediation at the 9-month follow-up (indirect effect = -.06; 95% CI [-.12, -.007]). Next, we tested the indirect effect of motivational interviewing on symptoms of somatization through self-efficacy regarding alcohol use for both follow-ups. We confirmed that motivational interviewing predicted a reduction in symptoms of somatization indirectly through self-efficacy regarding alcohol use at both follow-ups. The PROCESS macro found significant mediation with percentile bootstrap confidence intervals (CIs) of the indirect effect at the 3-month follow-up (indirect effect = -.05; 95% CI [-.12, -.009]) and at the 9-month follow-up (indirect effect = -.04; 95% CI [-.09, -.004]).

General Severity of Symptoms

The BSI-18 also produces a Global Severity Index that indicates the overall severity of symptomatology. Table 2 presents the means for general symptom severity by condition. An ANOVA test using repeated measures reveals that the general severity of symptoms decreased significantly over time for both conditions ($F = 25.27$, $df = 1$, $p < .001$). One-way ANOVA analyses show that there was not a significant effect of condition on general severity of symptoms at either the 3-month ($F < .00$, $df = 1$, $p = .99$) or the 9-month follow-up ($F = .08$, $df = 1$, $p = .77$). One-way ANOVA additionally revealed that self-efficacy regarding alcohol use predicted a significant decrease in general severity of symptoms at the 3-month follow-up ($F = 1.80$, $df = 1$, $p < .01$) and the 9-month follow-up ($F = 1.93$, $df = 1$, $p < .01$) and self-efficacy regarding contraception use predicted a weak marginal decrease in general severity of symptoms at the 9-month follow-up ($F = 1.41$, $df = 1$, $p = .12$).

Pearson correlational analyses revealed weak significant negative correlations between self-efficacy regarding all three behaviors with the general severity of symptoms at all three time points. Table 3 presents the correlation coefficients for these relationships. Based on these analyses regarding the general severity of symptoms, we tested the indirect effect of motivational interviewing on general severity of symptoms through self-efficacy regarding alcohol use at the 3-month and 9-month follow-ups and through self-efficacy regarding contraception use at the 9-month follow-up. We confirmed that motivational interviewing predicted a decrease in general severity of symptoms indirectly through self-efficacy regarding alcohol use at the 3-month and 9-month follow-up, but we did not find an indirect effect through self-efficacy regarding contraception use at the 9-month follow-up. The PROCESS macro found significant mediation with percentile bootstrap confidence intervals (CIs) of self-efficacy regarding alcohol use at the 3-month follow-up (indirect effect = -.08; 95% CI [-.16, -.02]) and at the 9-month follow-up (indirect effect = -.06; 95% CI [-.12, -.008]). The PROCESS macro did not find significant mediation of self-efficacy regarding contraception at the 9-month follow-up (indirect effect = -.01; 95% CI [-.04, .01]).

DISCUSSION

The purpose of the present study was to examine how motivational interviewing relates to self-efficacy and in turn how this leads to changes in depressive symptoms. Motivational interviewing was not found to predict a decrease in depressive symptoms compared to the control group at any time point, but it was found to increase self-efficacy regarding alcohol at the 3-month and 9-month follow-ups. Consistent with the hypothesized relationship, depressive symptoms were negatively correlated with self-efficacy regarding all three behaviors at all three measurement points. Our results offer some support for the hypothesized mediational model, which predicted that motivational interviewing would lead to lower depressive symptoms by increasing self-efficacy. Of the three behaviors for which self-efficacy

was measured, self-efficacy regarding alcohol use confirmed the mediational model at both post-intervention follow-ups, but self-efficacy regarding contraception use and self-efficacy regarding smoking did not confirm the model at either post-intervention follow-up.

MAIN FINDINGS

We found that motivational interviewing led to an increase in self-efficacy compared to the control group for alcohol use at both the 3-month and the 9-month follow-ups. However, both the control and experimental groups experienced significant increases in self-efficacy over time regarding both alcohol use and contraception use, as well as a marginal increase in self-efficacy regarding tobacco use. These findings indicate that both the experimental and the control group interventions were effective in increasing self-efficacy regarding alcohol use and contraception use, but with a smaller effect on self-efficacy regarding tobacco use. The control group received information about alcohol, contraception, and tobacco (if applicable) and were given resources to change their behaviors, so it makes sense that self-efficacy would increase even for the control group (Velasquez, Sternberg, Floyd, Parrish, Kowalchuk, Stephens, … Mullen, 2017). However, since criteria for participation in the parent study included heavy drinking and ineffective contraception use, this shift could partially be the result of a regression to the mean, or a return to more normative drinking and contraceptive habits. A unique effect of motivational interviewing in increasing self-efficacy compared to the control group was found only for self-efficacy regarding alcohol use. The effect of motivational interviewing on only one behavior could be due to the nature of the parent study, in which participants were encouraged to change one or more of the target behaviors, but not compelled to address them all. Figures 2, 3, and 4 show the relationship between self-efficacy and condition over time for alcohol use, contraception use, and tobacco use, respectively.

Just as self-efficacy generally increased for both groups, depressive symptoms significantly decreased over time for both the experimental and control groups. Both the control condition, which involved a brief advice session with a Behavioral Health specialist, and the experimental group, which involved two motivational interviewing sessions, experienced lower depressive symptoms. Since the initial averages for depressive symptoms were not considered clinically elevated (Derogatis, 1993), this result does not seem likely to represent regression to the mean. The present study did not find a direct effect of motivational interviewing on depressive symptoms. If the control group had received no form of treatment, rather than a brief advice session, we may have been able to identify the effects of motivational interviewing on depressive symptoms. The level of depressive symptoms that participants reported at baseline did not interact with the extent to which condition predicted changes in depressive symptoms. In other words, the effects of motivational interviewing were not strengthened for participants who initially reported either more or less depressive symptoms. This means that the treatment effect was general and not specific to any one level of initial depressive symptoms. Figure 5 represents the relationship between condition and depressive symptoms over time.

Although depressive symptoms were not significantly lower for the experimental condition compared to the control condition, they were significantly correlated with self-efficacy regarding all three behaviors at all time points. Higher self-efficacy predicted lower depressive symptoms for all three behaviors. This relationship, although significant, was relatively weak for all behaviors at all time points – with all correlation coefficients under .5 – but the strongest correlation was found between depressive symptoms and self-efficacy regarding alcohol use. These findings are consistent with the hypothesized relationship between self-efficacy and depressive symptoms.

The hypothesized mediational model was confirmed for self-efficacy regarding alcohol use at both time points. Although there was no main effect found for motivational interviewing on depressive symptoms, Hayes's PROCESS macro was able to find an indirect effect through bootstrapping (2018). The main effect was likely suppressed due to the opposing directions of the relationship between the independent variable and the mediator and between

the mediator and the dependent variable. Since motivational interviewing was positively correlated with self-efficacy, and self-efficacy was negatively correlated with depressive symptoms, the change in direction suppressed the main direct effect of motivational interviewing on depressive symptoms. With the PROCESS macro, however, motivational interviewing was found to indirectly decrease depressive symptoms by increasing self-efficacy regarding alcohol use. As mentioned earlier, the goal of the parent study was to reduce the risk of having an alcohol-exposed and/or tobacco-exposed pregnancy by changing one or more target behaviors, but not necessarily all of them, which may help partially explain why the mediational model was only confirmed for one of the three behaviors.

ADDITIONAL FINDINGS

The parent study used the BSI-18 to measure symptoms of psychological distress. The depression subscale was used for analysis of depressive symptoms, but the BSI-18 also includes subscales measuring symptoms of anxiety and somatization. When we analyzed symptoms of anxiety over time, we found that they mirrored depressive symptoms in their relationship to self-efficacy and motivational interviewing. Similar to depressive symptoms, symptoms of anxiety decreased over time across both conditions, and there was not a unique direct effect of motivational interviewing on symptoms of anxiety. Figure 6 presents changes in symptoms of anxiety over time by condition. Symptoms of anxiety were significantly negatively correlated with decreased self-efficacy for all three behaviors at all three time points with very similar correlation coefficients to those for self-efficacy and depressive symptoms (see Tables 1 and 2). Moreover, when we modified the hypothesized mediational model to replace depressive symptoms with symptoms of anxiety, we confirmed that motivational interviewing indirectly predicted a decrease in symptoms of anxiety by increasing self-efficacy regarding alcohol use at both follow-ups.

As for symptoms of somatization, they decreased over time across both conditions similarly to both depressive symptoms and symptoms of anxiety. In addition to decreasing linearly, symptoms of somatization also decreased quadratically, meaning they decreased sharply at the first follow-up and the effect was maintained over time. Figure 7 presents changes in symptoms of somatization over time by condition. However, unlike depressive symptoms and symptoms of anxiety, we found a weak marginal direct effect of motivational interviewing on symptoms of somatization, with motivational interviewing predicting lower symptoms of somatization. Symptoms of somatization were significantly correlated with lower self-efficacy for most measurement points. Self-efficacy regarding alcohol use significantly predicted lower symptoms of somatization at the 9-month follow up, and self-efficacy regarding contraception use weakly predicted lower symptoms of somatization at the 9-month follow-up. When we modified the hypothesized mediational model to replace depressive symptoms with symptoms of somatization, we confirmed that motivational interviewing indirectly predicted lower symptoms of somatization by increasing self-efficacy regarding alcohol use at both follow-ups.

Lastly, the BSI-18 produces a measure of general symptom severity called the General Severity Index (GSI) (Derogatis, 1993). Similar to the symptom subscales, general symptom severity decreased over time across both conditions, but there was not a unique direct effect of motivational interviewing on general symptom severity. In Figure 8, changes in general symptom severity over time by condition are presented. Also like the symptom subscales, higher self-efficacy was correlated with lower general symptom severity for all three behaviors at all three time points. Self-efficacy regarding alcohol use predicted lower general symptom severity at both follow-ups, and self-efficacy regarding alcohol use mediated the effect of motivational interviewing on general symptom severity at both follow-ups.

PRACTICAL IMPLICATIONS

As mentioned earlier, motivational interviewing is primarily used to treat behavioral-related problems, such as addictions and health behaviors. The present study suggests that, in addition to boosting one's confidence to resist, MI can also have positive effects on symptoms of anxiety, depression, and somatization. Since anxiety and depression are very often

comorbid with addiction and unhealthy behaviors (Lee, Oh, Cho, Hong, & Moon, 2001), MI can be extremely useful not only in changing unwanted behaviors but also in improving functionality and reducing distress for those who seek help with such problems.

Moreover, our findings regarding the correlates of self-efficacy are in line with previous research regarding the relationship between self-efficacy and mental health symptoms such as depression and anxiety (Faure & Loxton, 2003; Bandura, 1977). Previous research has also pointed to other beneficial correlates of self-efficacy, such as positive self-concept and skill acquisition and maintenance (Gist, Stevens, & Bavetta, 1991; Sagone & Elvira De Caroli, 2013). Since the present study confirmed that at least one kind of self-efficacy successfully mediated the relationship between MI and depressive symptoms, symptoms of anxiety, symptoms of somatization, and general symptom severity, it seems probable that MI could indirectly lead to other benefits through self-efficacy.

Additionally, the aforementioned exploratory research done by Khattra et al. (2017) pointed to the possibility of MI's utility for protecting against recurrence and achieving a more positive outlook for those with Generalized Anxiety Disorder when MI is added to traditional CBT treatment. The sample in the present study were not seeking treatment for cognitive-based disorders such as depression or anxiety, and the MI-based intervention did not involve any form of CBT. However, since the present study showed that MI indirectly reduced several different types of symptomatology by increasing self-efficacy, the addition of MI to a treatment plan seems promising for a broader array of psychological problems than those for which it is traditionally reserved.

LIMITATIONS

Limitations of the present study include its limited ability to assess the central constructs of the study due to the nature of the parent study. Though the parent study measured self-efficacy for two or three behaviors for each participant, self-efficacy was only measured in relation to the specified behaviors, rather than a more global representation of self-efficacy. As for depressive symptoms, the present study only included one measure of depressive symptoms – the BSI-18 depression subscale. In each of the three subscales of the BSI-18 (depression, anxiety, and somatization), there are only six items to measure the target construct, so this measure is not maximally robust.

The present study is also limited in generalizability due to the population sampled by the parent study. Participants did not include men, and they were largely Hispanic or non-Hispanic Black. A majority of participants were also low-income. Participants were selected for inclusion in the study because of heavy drinking and ineffective contraception use. These characteristics restrict the ability to assert the applicability of our findings to a more general population.

CONCLUSION

The present study partially confirmed the hypothesized mediational model, which predicted that MI would indirectly decrease depressive symptoms by increasing self-efficacy. The present study also found similar indirect effects of MI on symptoms of anxiety and somatization and on general symptom severity through self-efficacy. Future research should examine our mediational model with a more universal measure of self-efficacy and more robust measures of symptomatology and with a more general sample. The present study offers hopeful implications regarding MI's utility for improving a broad array of symptoms for those who are seeking to change a certain behavior and for those who are seeking treatment for problems that are not primarily behavior-based.

ACKNOWLEDGEMENTS

Special thanks to my faculty advisor, Kirk von Sternberg, Ph.D., for his guidance and giving me access to the data used for the present study. Thanks also to the Department of Social Work at the University of Texas at Austin, for conducting the parent study, and to the Departmental Honors Program in Psychology at the University of Texas at Austin, for guiding me through the process of research.

REFERENCES

Arkkukangas, M., Söderlund, A., Eriksson, S., & Johansson, A.-C. (2018). One-year adherence to the Otago exercise program with or without motivational interviewing in community-dwelling older adults. *Journal of Aging and Physical Activity, 26*(3), 390–395.

Arkowitz, H., Westra, H.A., Miller, W.R., & Rollnick, S. (2007). Motivational interviewing in the treatment of psychological problems. New York: Guilford Press.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215.

Berman, A. H., Forsberg, L., Durbeej, N., Källmén, H., & Hermansson, U. (2010). Single-session motivational interviewing for drug detoxification inpatients: effects on self-efficacy, stages of change and substance use. Substance Use & Misuse, 45, 384–402.

Borsari, B., Hopkins, L. B., Manuel, J. K., Apodaca, T. R., Mastroleo, N. R., Jackson, K. M., Magill, M., Norona, J. C., & Carey, K. B. (2019). Improvement in therapist skills over sessions in brief motivational interventions predicts client language and alcohol use outcomes. *Psychology of Addictive Behaviors, 33*(5), 484–494.

Breslin, F. C., Sobell, L. C., Sobell, M. B., & Agrawal, S. (2000). A comparison of a brief and long version of the Situational Confidence Questionnaire. *Behaviour Research and Therapy, 38*(12), 1211–1220. https://doi-org.ezproxy.lib.utexas.edu/10.1016/S0005-7967(99)00152-7

Burke, B. L., Arkowitz, H., & Menchola, M. (2003). The efficacy of motivational interviewing: A meta-analysis of controlled clinical trials. *Journal of Consulting and Clinical Psychology, 71*(5), 843–861.

D'Amico, E. J., Houck, J. M., Tucker, J. S., Ewing, B. A., & Pedersen, E. R. (2017). Group motivational interviewing for homeless young adults: Associations of change talk with substance use and sexual risk behavior. *Psychology of Addictive Behaviors, 31*(6), 688–698.

Derogatis L.R. (1993). BSI Brief Symptom Inventory: Administration, Scoring, and Procedure Manual, 4th ed., Minneapolis, MN: National Computer Systems.

DiClemente, C. C., Prochaska, J. O., & Gibertini, M. (1985). Self-efficacy and the stages of self-change of smoking. *Cognitive Therapy and Research, 9*(2), 181–200.

Faure, S., & Loxton, H. (2003). Anxiety, depression and self-efficacy levels of women undergoing first trimester abortion. *South African Journal of Psychology, 33*(1), 28–38. https://doi-org.ezproxy.lib.utexas.edu/10.1177/008124630303300104

Frost, H., Campbell, P., Maxwell, M., O'Carroll, R. E., Dombrowski, S. U., Williams, B., Cheyne, H., Coles, E., & Pollock, A. (2018). Effectiveness of Motivational Interviewing on adult behaviour change in health and social care settings: A systematic review of reviews. *PLoS ONE, 13*(10).

Gist, M. E., Stevens, C. K., & Bavetta, A. G. (1991). Effects of self-efficacy and post-training intervention on the acquisition and maintenance of complex interpersonal skills. *Personnel psychology, 44*(4), 837-861.

Hasin, D. S., Sarvet, A. L., Meyers, J. L., Saha, T. D., Ruan, W. J., Stohl, M., & Grant, B. F. (2018). Epidemiology of Adult DSM-5 Major Depressive Disorder and Its Specifiers in the United States. *JAMA Psychiatry, 75*(4), 336.

Hayes, A. F. (2018). *Introduction to Mediation, Moderation, and Conditional Process Analysis* (2nd ed.). New York: Guilford.

Hollon, S.D., Jarrett, R.B., Nierenberg, A.A., Thase, M.E., Trivedi, M., & Rush, A.J. (2005). Psychotherapy and medication in the treatment of adult and geriatric depression: which monotherapy or combined treatment? Journal of Clinical Psychiatry, 66, 455-468.

Honyashiki, M., Furukawa, T. A., Noma, H., Tanaka, S., Chen, P., Ichikawa, K., Ono, M., Churchill, R., Hunot, V., & Caldwell, D. M. (2014). Specificity of CBT for depression: A contribution from multiple treatments meta-analyses. *Cognitive Therapy and Research, 38*(3), 249–260.

Huang, F.-F., Yang, Q., Zhang, J., Han, X. Y., Zhang, J.-P., & Ye, M. (2018). A self-efficacy enhancing intervention for pulmonary rehabilitation based on motivational interviewing for postoperative lung cancers patients: Modeling and randomized exploratory trial. *Psychology, Health & Medicine, 23*(7), 804–822.

Keeley, R. D., Brody, D. S., Engel, M., Burke, B. L., Nordstrom, K., Moralez, E., Dickinson, L. M., & Emsermann, C. (2016). Motivational interviewing improves depression outcome in primary care: A cluster randomized trial. *Journal of Consulting and Clinical Psychology, 84*(11), 993–1007.

Khattra, J., Angus, L., Westra, H., Macaulay, C., Moertl, K., & Constantino, M. (2017). Client perceptions of corrective experiences in cognitive behavioral therapy and motivational interviewing for generalized anxiety disorder: An exploratory pilot study. *Journal of Psychotherapy Integration, 27*(1), 23–34.

Lee, M. S., Oh, E. Y., Cho, S. M., Hong, M. J., & Moon, J. S. (2001). An assessment of adolescent Internet addiction problems related to depression, social anxiety and peer relationship. *Journal of Korean Neuropsychiatric Association, 40*(4), 616-628.

Miller, W. R., & Moyers, T. B. (2017). Motivational interviewing and the clinical science of Carl Rogers. *Journal of Consulting and Clinical Psychology, 85*(8), 757–766.

Miller, W. R. & Rollnick, S. (2002). Motivational interviewing: Preparing people for change. New York: Guilford Press.

Miller, P. J., Ross, S. M., Emmerson, R. Y., & Todt, E. H. (1989). Self-efficacy in alcoholics: Clinical validation of the Situational Confidence Questionnaire. *Addictive Behaviors, 14*(2), 217–224. https://doi-org.ezproxy.lib.utexas.edu/10.1016/0306-4603(89)90052-X

Moyers, T. B., Martin, T., Manuel, J. K., Miller, W. R., & Ernst, D. (2010). *Revised global scales: motivational interviewing treatment integrity 3.1.1.* Center on Alcoholism, Substance Abuse and Addictions.

Muir, H. J., Constantino, M. J., Coyne, A. E., Westra, H. A., & Antony, M. M. (2019). Integrating responsive motivational interviewing with cognitive–behavioral therapy (CBT) for generalized anxiety disorder: Direct and indirect effects on interpersonal outcomes. *Journal of Psychotherapy Integration*.

Pierson, H. M., Hayes, S. C., Gifford, E. V., Roget, N., Padilla, M., Bissett, R., Berry, K., Kohlenberg, B., Rhode, R., & Fisher, G. (2007). An examination of the Motivational Interviewing Treatment Integrity code. *Journal of Substance Abuse Treatment, 32*(1), 11–17.

Pfund, R. A., Whelan, J. P., Peter, S. C., & Meyers, A. W. (2020). Can a motivational letter increase attendance to psychological treatment for gambling disorder? *Psychological Services, 17*(1), 102–109.

Sagone, E. & Elvira De Caroli, M.. (2013). Locus of control and academic self-efficacy in university students: the effects of self-concepts. *Procedia-Social and Behavioral Sciences, 114*, 222-228.

Velasquez, M. M., Sternberg, K. L. V., Floyd, R. L., Parrish, D., Kowalchuk, A., Stephens, N. S., … Mullen, P. D. (2017). Preventing Alcohol and Tobacco Exposed Pregnancies: CHOICES Plus in Primary Care. *American Journal of Preventive Medicine, 53*(1), 85–95.

Westra, H.A., Arkowitz, H., & Dozois, D.A. (2009). Adding a motivational interviewing pretreatment to cognitive behavioral therapy for generalized anxiety disorder: A preliminary randomized controlled trial. Journal of Anxiety Disorders, 23, 1106- 1117.

Westra, H. A., & Aviram, A. (2013). Core skills in motivational interviewing. *Psychotherapy, 50*(3), 273–278.

Worley, M. J., Tate, S. R., Granholm, E., & Brown, S. A. (2014). Mediated and moderated effects of neurocognitive impairment on outcomes of treatment for substance dependence and major depression. *Journal of Consulting and Clinical Psychology, 82*(3), 418–428.

Self-efficacy				
		Alcohol Use	Contraception Use	Tobacco Use
Depressive Symptoms	Baseline	-.42**	-.17**	-.21*
	3 months	-.25**	-.16*	-.45**
	9 months	-.34**	-.14*	-.32**
				*p < .05, **p < .01

Table 1: Pearson correlations between self-efficacy regarding the three behaviors and depressive symptoms at the three measurement points.

Condition			
		Motivational Interviewing	Control
Symptoms of Anxiety	Baseline	.89	.82
	3 months	.72	.63
	9 months	.61	.59
Symptoms of Somatization	Baseline	.73	.76
	3 months	.54	.69
	9 months	.61	.69
General Symptom Severity	Baseline	.90	.85
	3 months	.70	.70
	9 months	.64	.66

Table 2: Means for measures of symptoms of anxiety, symptoms of somatization, and general symptom severity by condition at the three measurement points.

Self-efficacy				
		Alcohol Use	Contraception Use	Tobacco Use
Symptoms of Anxiety	Baseline	-.35**	-.17**	-.20**
	3 months	-.21**	-.13*	-.43**
	9 months	-.28**	-.10	-.34**
Symptoms of Somatization	Baseline	-.28**	-.14*	-.04
	3 months	-.15*	-.11	-.23*
	9 months	-.23**	-.10	-.28**
General Symptom Severity	Baseline	-.40**	-.18**	-.18*
	3 months	-.29**	-.19**	-.41**
	9 months	-.32**	-.13*	-.34**
				*p < .05, **p < .01

Table 3: Pearson correlations between self-efficacy regarding the three behaviors and symptoms of anxiety, symptoms of somatization, and general symptom severity at the three measurement points.

Figure 1. A graphic representation of the hypothesized mediational model.

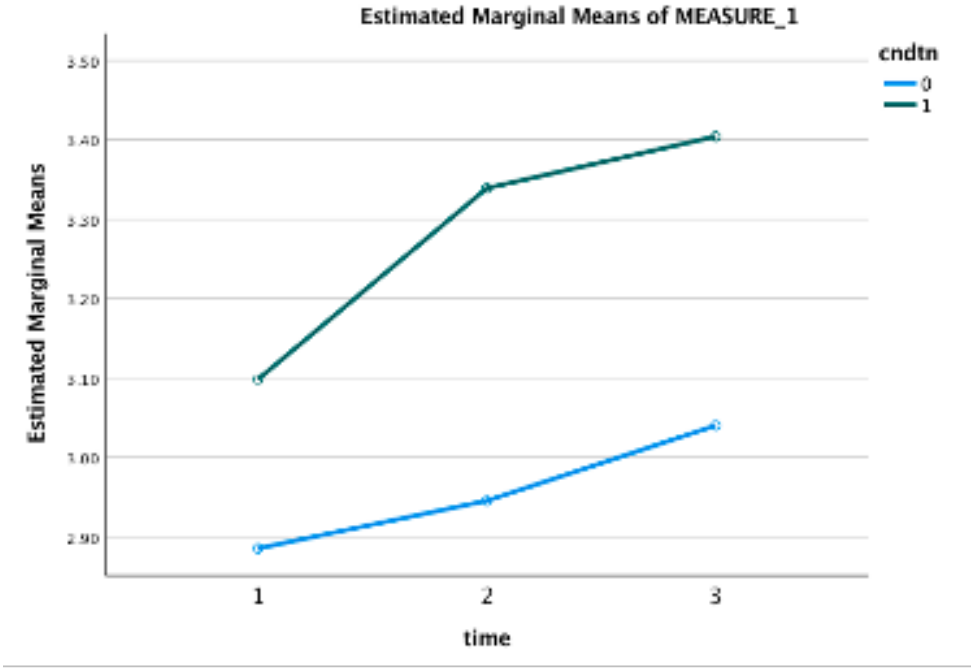


Figure 2. Self-efficacy regarding alcohol use over time by condition. For condition, 0 indicates the control group, and 1 indicates the experimental (MI) group. For time, 0 represents the pre-intervention baseline, 1 represents the 3-month-follow-up, and 3 represents the 9-month follow-up.

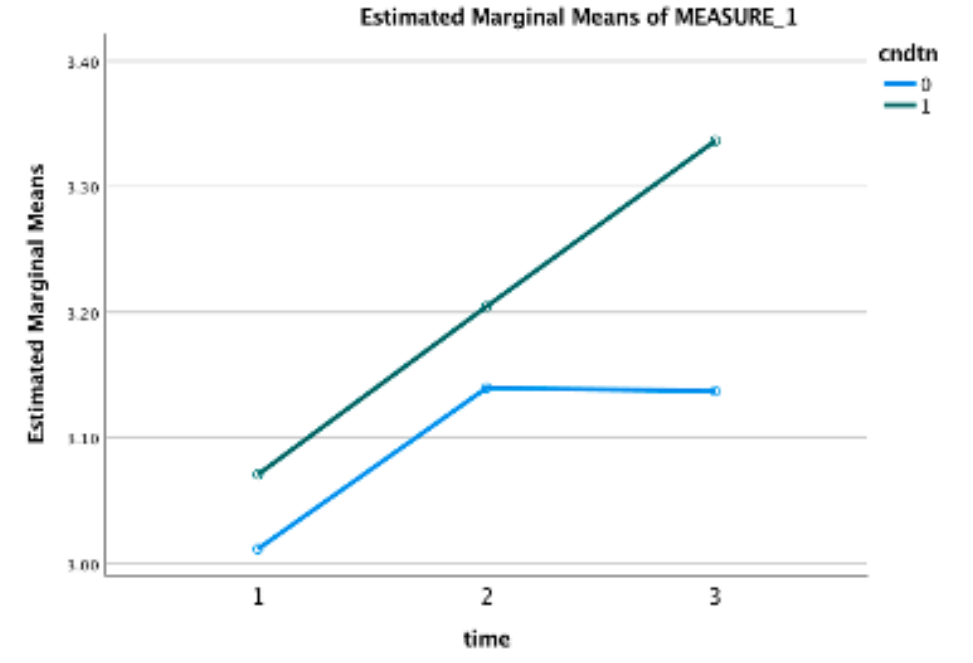


Figure 3. Self-efficacy regarding contraception use over time by condition. For condition, 0 indicates the control group, and 1 indicates the experimental (MI) group. For time, 0 represents the pre-intervention baseline, 1 represents the 3-month-follow-up, and 3 represents the 9-month follow-up.

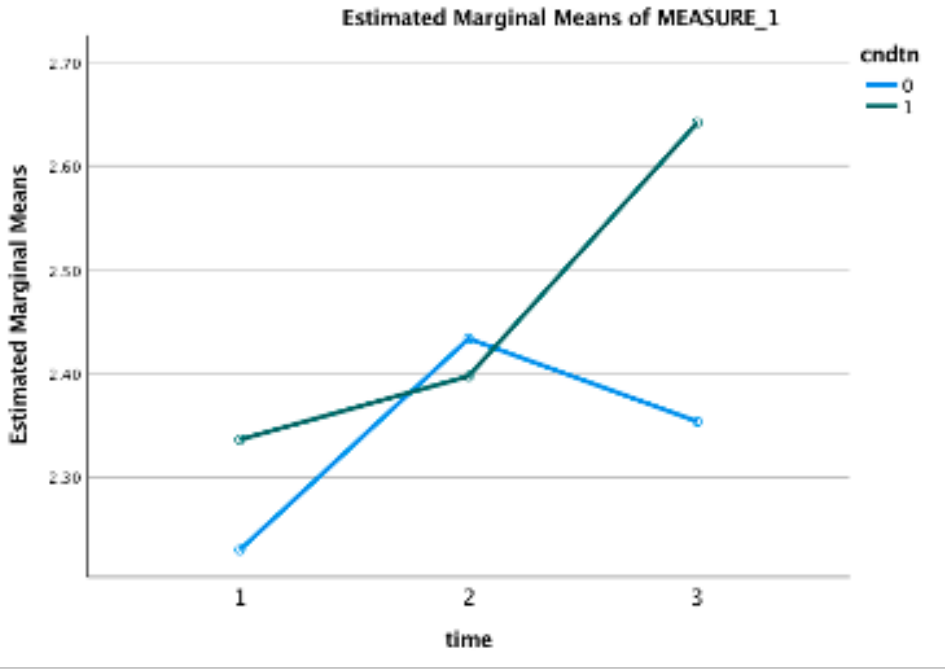


Figure 4. Self-efficacy regarding tobacco use over time by condition. For condition, 0 indicates the control group, and 1 indicates the experimental (MI) group. For time, 0 represents the pre-intervention baseline, 1 represents the 3-month-follow-up, and 3 represents the 9-month follow-up.

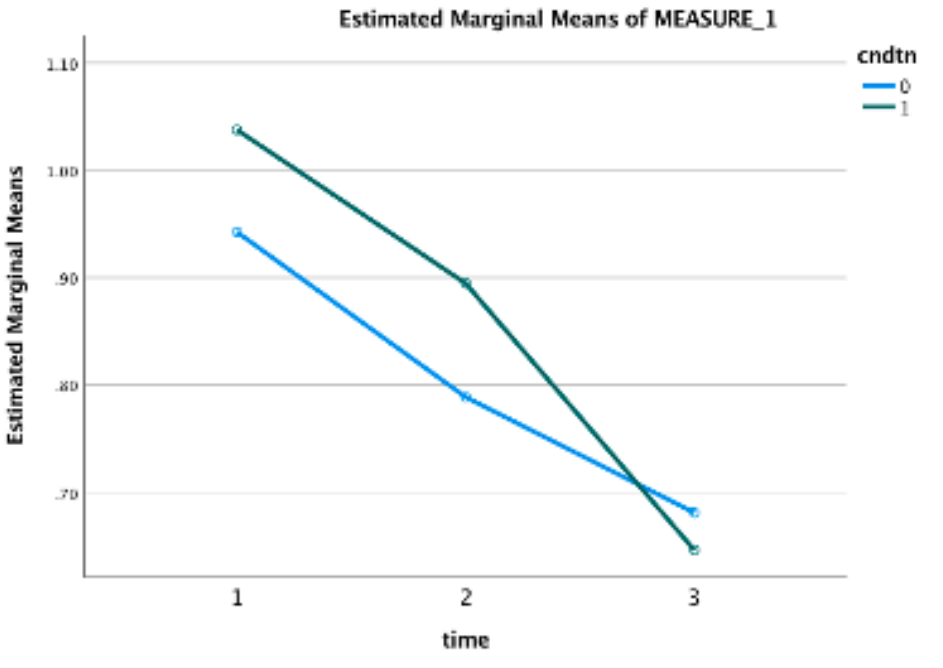


Figure 5. Depressive symptoms over time by condition. For condition, 0 indicates the control group, and 1 indicates the experimental (MI) group. For time, 0 represents the pre-intervention baseline, 1 represents the 3-month-follow-up, and 3 represents the 9-month follow-up.

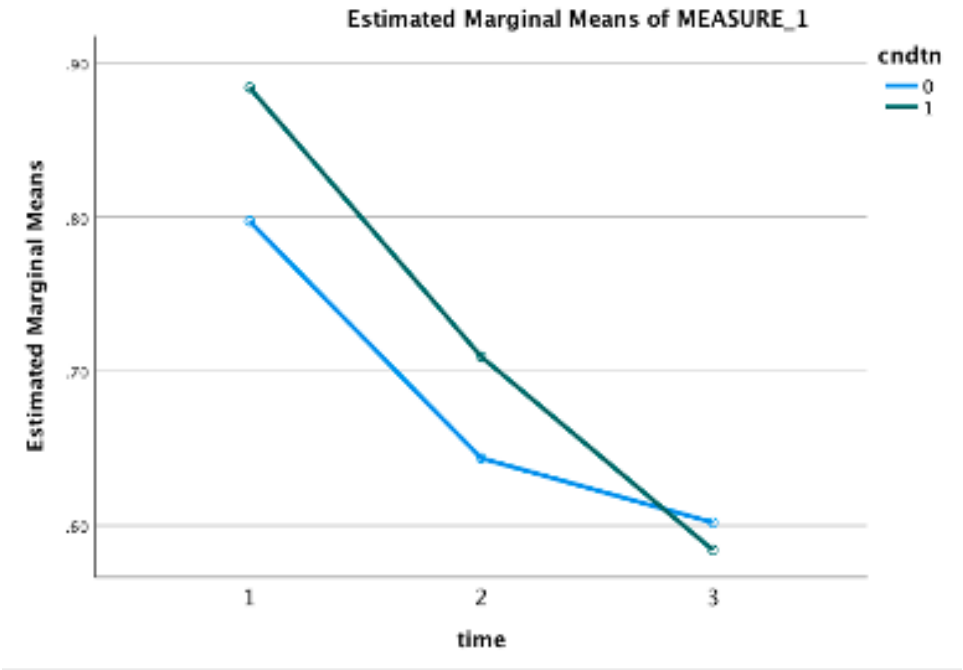


Figure 6. Symptoms of anxiety over time by condition. For condition, 0 indicates the control group, and 1 indicates the experimental (MI) group. For time, 0 represents the pre-intervention baseline, 1 represents the 3-month-follow-up, and 3 represents the 9-month follow-up.

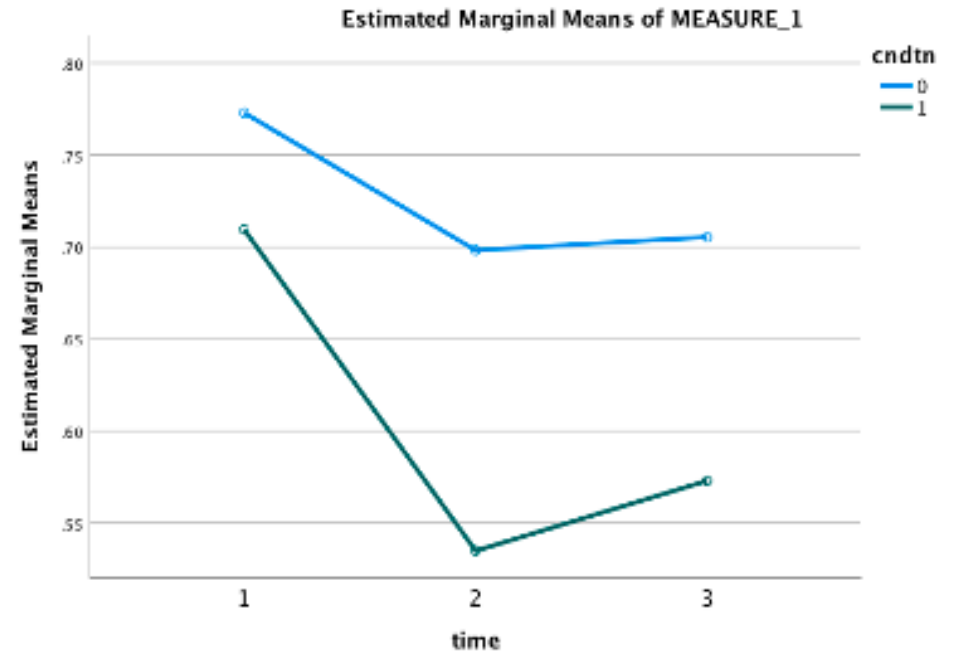


Figure 7. Symptoms of somatization over time by condition. For condition, 0 indicates the control group, and 1 indicates the experimental (MI) group. For time, 0 represents the pre-intervention baseline, 1 represents the 3-month-follow-up, and 3 represents the 9-month follow-up.

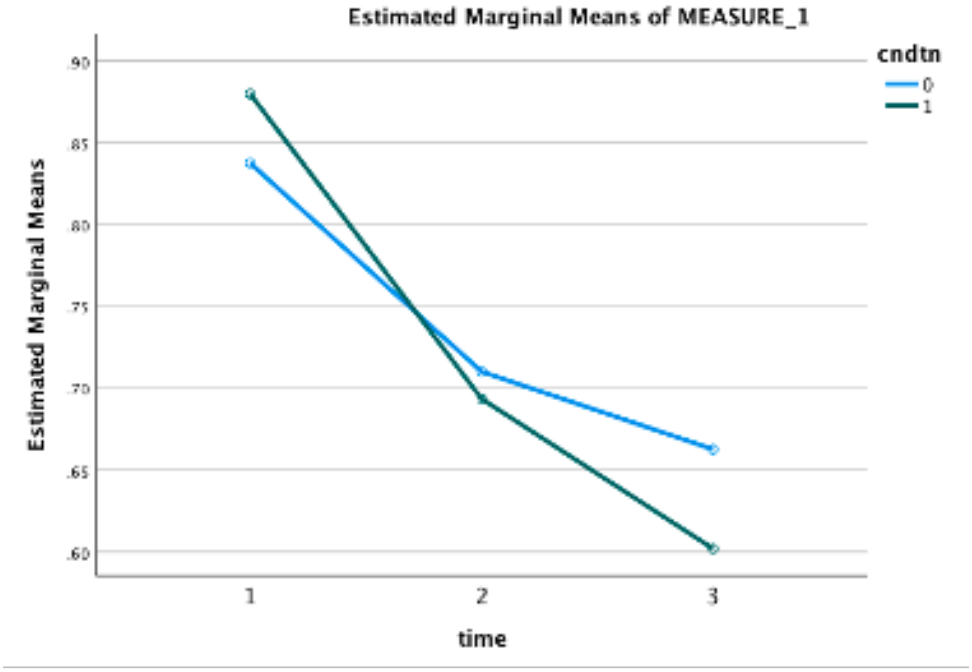


Figure 8. General symptom severity over time by condition. For condition, 0 indicates the control group, and 1 indicates the experimental (MI) group. For time, 0 represents the pre-intervention baseline, 1 represents the 3-month-follow-up, and 3 represents the 9-month follow-up.

Depression and Alcohol Use in Young Adults during the COVID-19 Pandemic and relations with Insula Structure

Raquel Kosted
Department of Neuroscience, The University of Texas at Austin
Texas Undergraduate Research Journal

ABSTRACT

Recent studies have shown that the COVID-19 pandemic has resulted in an unprecedented increase in stress and mental health issues globally. The underlying neurobiology that may contribute to differences in COVID-related stress, and associated mood and alcohol use outcomes is unclear. This study investigates the relationship between COVID-related stress and change in depressed mood from pre-COVID-19 pandemic to during the COVID-19 pandemic in young adults with a mood disorder (major depressive disorder and bipolar disorder), compared to typically developing young adults. We further investigated how change in depression related to change in alcohol use from pre-pandemic to during the pandemic. Additionally, we investigated how insula structure assessed at baseline (pre-COVID-19 pandemic) related to COVID-related stress and change in depressed mood and alcohol use at the COVID follow-up assessment. We found that greater COVID-related stress was associated with greater increase in depression symptoms during the pandemic, and greater increase in depression symptoms was associated with lower left insula cortical thickness in all participants. In young adults with a mood disorder, greater increase in depression symptoms was associated with greater increase in alcohol use during the pandemic, and greater increase in alcohol use was associated with lower left insula cortical thickness. These results suggest that lower left insula cortical thickness relates to greater depression symptoms in all young adults, and greater alcohol use in young adults with a mood disorder, during the pandemic. Additionally, our results suggest greater increase in depression symptoms relate to greater increase in alcohol use in young adults with a mood disorder during the COVID-19 pandemic.

BACKGROUND

The COVID-19 pandemic is a tremendous global health crisis, with more than 136 million cases and 2.9 million deaths worldwide as of April 14, 2021 [1]. Recent studies have shown this pandemic has resulted in an unprecedented increase in stress and mental health issues globally [2]. Pandemic-related stressors include perception of safety, threat and risk of contagion, as well as quarantine and confinement [3, 4, 5]. The prevalence of depression and alcohol use has increased during the COVID-19 pandemic [6, 7]. Studies have shown that individuals with mood disorders exhibit greater COVID-related stress [8] and are at greater risk for depression and alcohol use during the COVID-19 pandemic compared to the general public [9, 10].

Individuals with mood disorders show increased sensitivity to stress [11, 12]. Dysregulation of stress pathways in the brain has been associated with severity of clinical symptoms (i.e. depressive symptoms) [13, 14, 15] and maladaptive coping strategies (such as alcohol misuse) during stress [16, 17]. Therefore, it is possible that greater COVID-related stress may be one factor contributing to the higher rate of depression and alcohol use observed in individuals with a mood disorder during the pandemic. The

relationship between depressed mood and alcohol use [18] should be considered when interpreting the interactions of stress, depression, and alcohol use.

Mood disorders, including major depressive disorder and bipolar disorder, commonly co-occur with alcohol use disorder [19]. Using alcohol to reduce emotional distress (self-medication) has been proposed as one explanation for the high comorbidity of these disorders [20]. Indeed, research has shown that individuals with mood disorders exhibit difficulties with emotional regulation, and may engage in heavier drinking to cope with negative affect [21]. Therefore, individuals with mood disorders may be at greater risk for increasing alcohol use as a means to cope with adverse mood symptoms experienced during the pandemic. In fact, a recent study found that individuals with severe depressive symptoms were more likely to increase alcohol use during the pandemic than the general public [22]. While recent studies have investigated changes in mood and alcohol use across individuals during the COVID-19 pandemic, few studies have investigated the underlying neurobiology that may predict the degree to which an individual is stressed by the pandemic, and how much their mood and alcohol use will change.

The underlying neurobiology that may contribute to COVID-related stress, and associated mood and alcohol use outcomes is unclear. Hyperactivity of the insula has been associated with heightened stress response [23], and differences in insula structure have been associated with depression and alcohol use disorder [24, 25]. The insula appears to play an important role in subjective feeling states [26], and it has been hypothesized that dysfunction of the insula may contribute to the negative mood states experienced by individuals with depression [27, 28]. Within individuals with bipolar disorder, lower insula grey matter volume is associated with increased risk for alcohol use problems [29]. Since the insula has been implicated to play a role in the regulation of stress, depression, and alcohol use, insula structure may predict individuals' stress response to COVID-19, and associated changes in mood and alcohol use during the pandemic.

The current study investigated relations between COVID-related stress and changes in depressed mood from before to during the COVID-19 pandemic in young adults with a mood disorder (major depressive disorder and bipolar disorder), compared to typically developing young adults, and studied how changes in depression relates to changes in alcohol use from before the pandemic to during the pandemic. We further investigated how insula structure assessed at the baseline (pre-COVID-19 pandemic) relates to COVID-related stress, and changes in depressed mood and alcohol use at the COVID follow-up assessment. Participants completed an assessment of recent depressive symptoms, alcohol use, and structural magnetic resonance imaging (sMRI) to assess insula cortical thickness prior to the COVID-19 pandemic (the baseline assessment). During the COVID-19 pandemic, participants filled out a survey on stressful experiences related to the pandemic and repeated the same assessments of recent depressive symptoms and alcohol use. Changes in depression and alcohol use from before the pandemic to during the pandemic were calculated for each participant by subtracting pre-pandemic scores from during-pandemic scores. We hypothesized that young adults with a mood disorder would exhibit greater COVID-related stress, and a greater increase in depressive symptoms and alcohol use. We predicted that greater COVID-related stress would relate to greater increases in depression and alcohol use in both groups. We also predicted that greater increase in depression would relate to greater increase in alcohol use. We further predicted that lower insula thickness would relate to greater COVID-related stress and greater increases in depression and alcohol use during the COVID-19 pandemic.

METHODS

Participants and Data Collection

The current study included 37 young adults (19 participants with a mood disorder (major depressive disorder and bipolar disorder) and 18 healthy comparison participants; see Ttable 1 for demographic and clinical characteristics). All of the participants completed a baseline assessment (pre-COVID-19) and a follow-up assessment during the COVID-19 pandemic, which occurred on average 0.77 years

following the baseline assessment. A subset of participants (6 with a mood disorder and 9 typically developing) completed a MRI scan at the baseline assessment. The Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders 5 (DSM-5) Research Version (SCID-5-RV) [30] was conducted to assess for diagnosis of bipolar disorder, major depressive disorder, anxiety disorders, and alcohol/substance use disorders at the baseline assessment. Verbal comprehension and matrix reasoning subtests of the Wechsler Abbreviated Scale of Intelligence-Second Edition (WASI-II) were used as a measure of full-scale intelligence quotient (FSIQ-2). For all of the participants, exclusion criteria included a history of any major medical illness with possible neurological or central nervous system outcomes, IQ < 85, being pregnant, or having a medical condition/previous surgery preventing participation in a MRI scan. Exclusion criteria for typically developing participants included a history of mood episodes, psychosis, or a lifetime suicide attempt, and use of psychotropic medication. Urinalysis was conducted on the day of the scan to assess for pregnancy and substance use. Participants were asked to abstain from any alcohol or drug use for 24-hours preceding the MRI scan. All study procedures were approved by the University of Texas at Austin Institutional Review Board and written consent was obtained from all of the participants.

Assessment of Mood, Alcohol Use, and COVID-related Stress

The Beck Depression Inventory [31] was used to assess current depression symptoms at the baseline and again at COVID follow-up assessments (see table 3). Recent alcohol use was assessed at the baseline and COVID follow-up assessments using the Daily Drinking Questionnaire (DDQ), modified for heaviest drinking week (DDQ-H) over the past month [32]. The total number of drinks consumed and the total number of drinking days during the heaviest drinking week were calculated (see Ttable 4). The adult self-report baseline short form version of the Coronavirus Health Impact Survey V0.2 [33] was used to assess COVID-related stress at the COVID follow-up assessment. This survey collected measures of participants' worry of being infected, how much COVID-19 has affected mental/emotional health, stress of restrictions on leaving home, difficulty following the recommendations for keeping away from close contact with people, stress of changes in family contacts, stress of changes in social contacts, and hopefulness that the COVID-19 crisis will end soon. Scores for each of these stress-related items were summed to calculate a COVID-related total stress score for each participant.

MRI Acquisition and Preprocessing

Magnetic Resonance Imaging was performed with a single 3-Tesla Siemens Skyra MR scanner using a 32-channel head coil located at the University of Texas at Austin Biomedical Imaging Center. Sagittal structural MRI images were acquired with a three-dimensional MPRAGE T1-weighted sequence with the following parameters: repetition time (TR) = 1900ms, echo time (TE) = 2.42 ms, matrix = 224 x 224, field of view = 220 x 220 mm2, 192 one-mm slices without gap and one average. All of the scans were assessed visually for movement and noise artifacts. FreeSurfer version 7.1 (https://surfer.nmr.mgh.harvard.edu) was used for cortical surface reconstruction and to obtain a measure of insula cortical thickness (www.enigma.ini.usc.edu). The insula region of interest was defined using the Desikan-Killiany atlas [34]. The measure of cortical thickness for the region of interest was calculated in FreeSurfer and extracted for subsequent analyses.

Between Group Differences in Demographics and Clinical Factors

Between group differences in the continuous demographic and clinical variables were assessed with two-sample t-tests, and included age and years between the baseline and COVID follow-up assessments. Between group differences in categorical variables were assessed with Chi square or Fisher's exact tests, as appropriate, and included sex, race, past and current alcohol use disorders, past and current cannabis use disorders, and anxiety disorders. Within the mood disorder group, percent of participants with major depressive disorder, bipolar disorder, rapid cycling, lifetime psychosis, and lifetime suicide attempt was calculated. Between group differences in mood state on the day of MRI scan

was assessed for participants in the neuroimaging data set with Fisher's exact tests.

Between group differences in the continuous demographic and clinical variables were assessed with two-sample t-tests, and included age and years between the baseline and COVID follow-up assessments. For categorical variables, between group differences were assessed with Chi square or Fisher's exact tests, as appropriate, and included sex, race, past and current alcohol use disorders, past and current cannabis use disorders, and anxiety disorders. Within the mood disorder group, percents of participants with major depressive disorder, bipolar disorder, rapid cycling, lifetime psychosis, and lifetime suicide attempt were calculated. Variations between groups in mood state on the day of MRI scan were also assessed for participants in the neuroimaging data set with Fisher's exact tests.

Between Group Differences in COVID-Related Stress, Changes in Depression and Alcohol Use

Change in depression from pre-COVID to during-COVID was calculated for each participant (given by the Beck Depression Inventory score at COVID follow-up minus the Beck Depression Inventory score at baseline assessment). Similarly, change in total number of drinks during heaviest drinking week and total number of drinking days during heaviest drinking week from pre-COVID to during-COVID was calculated for each participant. Between group differences in COVID-related stress, change in depression, and change in alcohol use were calculated using t-tests. Between group differences in depression and alcohol use at baseline assessment, as well as at COVID follow-up, were calculated using t-tests.

Relations between COVID-Related Stress, Changes in Depression and Alcohol Use

The main effect of total COVID-related stress, group, and total COVID-related stress by group interactions on change in depression (dependent variable) was modeled using a Standard Least Squares Fit Model, with biological sex included as a covariate. Parallel models were conducted to assess the relation between total COVID-related stress, group, and total COVID-related stress by group interactions on change in alcohol use (modeled separately for total number of drinks and for number of drinking days). Additionally, Parallel models were conducted to assess relations between change in depression, group, and change in depression by group interactions on change in alcohol use (modeled separately for total number of drinks and for number of drinking days). All models were repeated stratified by group to explore relations within each group. Significance was defined as alpha < 0.05 for these planned analyses.

Relations between Insula Structure and COVID-Related Stress and Changes in Depression Symptoms and Alcohol Use

Main effect of insula thickness, group, and insula thickness by group interactions on COVID-related stress (dependent variable) was modeled using a Standard Least Squares Fit Model, with biological sex included as a covariate. Separate models were run for left and right insula. Parallel models were conducted to assess relations between insula thickness, group, and insula thickness by group interactions and change in depression and change in alcohol use (modeled separately). Models were repeated stratified by group to explore relations within each group. Significance was defined as alpha < 0.05 for these planned analyses. Significant results are reported below.

RESULTS

Demographic and Clinical Factor Analyses

The mood disorder group included a greater number of participants with an anxiety disorder ($p = 0.02$). No other between group differences was observed (see table 1).

Between Group Differences in COVID-Related Stress, Depression, and Alcohol Use

At the COVID follow-up, the mood disorder group showed significantly higher Beck Depression Inventory scores ($p = 0.005$) compared to the healthy group, but groups did not differ in alcohol use (see table 3). Groups did not differ in depression or alcohol use at the pre-COVID assessment. Groups also did not differ in change in depression or change in alcohol use pre-COVID to during-COVID (see table 4). There was no difference between groups in total COVID-related stress scores.

Relations between COVID-Related Stress, Depression, and Alcohol Use

Greater total COVID-related stress was associated with greater increase in Beck Depression score (change from pre-COVID to during the COVID-19 pandemic) across all participants (main effect of COVID-related stress, $p = 0.002$). Stratifying by group revealed positive relation between COVID-related stress and change in depression within both the mood disorder group ($p = 0.03$) and the healthy comparison group ($p = 0.04$) (figure 1). Total COVID-related stress was not significantly related to increase in alcohol use (change from pre-COVID to during the COVID-19 pandemic) across participants.

Greater increase in Beck Depression score from pre-COVID to during-COVID was associated with greater increase in total number of drinks consumed during heaviest drinking week from pre-COVID to during-COVID (main effect of change in depression score, $p = 0.03$). Stratifying by group revealed a positive relation between change in depression and change in drinking within the mood disorder group ($p = 0.04$, significant), and no significant relation between change in depression and change in drinking within the healthy comparison group ($p = 0.63$) (figure 2).

Insula Structure and COVID-Related Stress, Depression, and Alcohol Use

There was a significant main effect of left insula thickness on change in Beck Depression score (change from pre-COVID to during the COVID-19 pandemic) ($p = 0.002$) such that lower measures of left insula thickness were associated with greater increases in depression during the pandemic. Stratifying by group revealed negative relation between insula thickness and change in depression within the healthy comparison group ($p = 0.04$, significant), and no significant relation between insula thickness and change in depression within the mood disorder group ($p = 0.09$) (figure 3).

There was a significant main effect of left insula thickness on change in total drinks consumed in heaviest drinking week (change from pre-COVID to during the COVID-19 pandemic) ($p = 0.008$) such that lower measures of left insula thickness were associated with greater increases in alcohol use during the pandemic. There was a significant group by left insula thickness interaction on change in total number of drinks consumed in heaviest drinking week ($p = 0.0038$). Stratifying by group revealed a significant negative relation between left insula thickness and change in total number of drinks consumed in heaviest drinking week in the mood disorder group ($p = 0.002$), but no relation in the healthy control group ($p = 0.28$) (figure 4). Specifically, those in the mood disorder group with lower insula thickness showed greater increases in quantity of alcohol use during the pandemic.

DISCUSSION

The current study was designed to assess relations between COVID-related stress, change in depressed mood, and change in alcohol use during the pandemic (compared to a recent pre-pandemic assessment) in young adults with a mood disorder (major depressive disorder or bipolar disorder) compared to typically developing young adults. Our study also assessed how differences in insula structure relate to COVID-related stress and change in depression and alcohol use during the COVID-19 pandemic. We found that greater COVID-related stress was associated with greater increase in depression symptoms

during the pandemic, and greater increase in depression symptoms was associated with lower left insula thickness across all participants. Within the mood disorder group, greater increase in depression symptoms was associated with greater increase in alcohol use during the pandemic, and greater increase in alcohol use was associated with lower left insula thickness.

Our results do not support our hypothesis that young adults with a mood disorder would exhibit greater COVID-related stress, and greater increase in depressive symptoms and alcohol use. The lack of a significant finding may be due to the diversity of the mood disorder group. Research has shown there are differences in whole-brain activity between individuals with major depressive disorder and bipolar disorder when exposed to emotion-provoking stimuli [35], which may contribute to differences in emotional stress reactivity [36]. Differences in response to the COVID-19 pandemic between individuals with different disorders within the mood disorder group may have made the data too noisy to produce a significant finding.

Our results support our hypothesis that COVID-related stress may be associated with increased depression symptoms in both young adults with a mood disorder and typically developing young adults during the pandemic. These results support previous research findings that suggest a relationship between heightened sensitivity to stress and subsequent depressive symptoms [14]. Our results do not support our hypothesis that COVID-related stress may be associated with increased alcohol use during the COVID-19 pandemic.

Our results support our hypothesis that increased depression symptoms may be associated with increased alcohol use during the COVID-19 pandemic. Although there was not a significant group by change in depression interaction on change in alcohol use, stratifying results by group showed that the mood disorder group is driving this finding. Increased depression symptoms are significantly related to increased alcohol use in young adults with a mood disorder, but not in typically developing young adults. These results are in line with previous research findings that suggest individuals with mood disorders struggle with self-regulation of emotion, and may use alcohol as an external remedy to reduce emotional distress [20, 21].

Our results do not support our prediction that lower insula thickness may be associated with greater COVID-related stress. However, the results of our study do suggest lower left insula thickness relates to a subsequent increase in depressed mood during the COVID-19 pandemic in both young adults with a mood disorder and typically developing young adults, as predicted. These results support the findings of previous research that suggest there is an association between insula structure and risk for development of symptoms of depression [24]. The results of our study also suggest that lower left insula thickness relates to greater increase in alcohol use during the COVID-19 pandemic, as predicted. These results support the findings of previous research that suggest there is an association between insula structure and risk for alcohol use problems [29]. Interestingly, after stratifying these results by group, it was revealed that the mood disorder group was driving this finding. Those in the mood disorder group with lower insula thickness showed greater increases in quantity of alcohol use during the pandemic, but this association does not appear in the healthy comparison group.

Relationships between change in depression symptoms and change in alcohol use, and between insula thickness and change in alcohol use, appear to be exhibited only by the mood disorder group. Increased alcohol use within the mood disorder group may be a maladaptive coping mechanism related to increased depression during the pandemic. Neurobiological differences, such as differences in insula structure, between young adults with a mood disorder and typically developing young adults may contribute to increased likelihood to use alcohol as a coping mechanism in negative affect situations within the mood disorder group.

There are some limitations to this study that should be discussed. The sample size for the behavioral/clinical dataset is relatively small (n = 37), and the sample size for the neuroimaging dataset is even smaller (n = 15). Due to small sample size, we were underpowered to explore differences between par-

ticipants with major depressive disorder and bipolar disorder. Data was collected during different points of the pandemic and participants' stress, depression, and alcohol use might have changed over the course of the pandemic. Stress, depression, and alcohol use measures were self-reported. An objective measure of stress (such as cortisol levels) was not collected. Differences in stressful experiences such as unemployment, financial problems, food insecurity and lack of health insurance were not looked at. Differences between groups that were not considered during data analysis could contribute to findings or lack thereof. The dataset may be biased since it only includes participants that agreed to do a COVID follow-up assessment. Participants that did not do the COVID follow-up assessment may have opted out because they were experiencing greater stress and depressed mood. Future studies with larger and more generalizable samples may see more robust differences in responses between young adults with a mood disorder and typically developing young adults.

ACKNOWLEDGEMENTS

Principal Investigator Supervisor:

Dr. Elizabeth Lippard, Dept. of Psychiatry at Dell Medical School

Graduate Student Supervisor:

Dylan Kirsch, Institute for Neuroscience at UT Austin

Lab Manager Supervisor:

Vanessa Le, Lab Manager of Dr. Lippard's lab, Dept. of Psychiatry at Dell Medical School

REFERENCES

1. "WHO Coronavirus Diseases (COVID-19) Dashboard." World Health Organization, *World Health Organization*, 30 Nov. 2020, covid19.who.int/.

2. Xiong, Jiaqi, et al. "Impact of COVID-19 Pandemic on Mental Health in the General Population: A Systemic Review." *Journal of Affective Disorders*, U.S. National Library of Medicine, 8 Aug. 2020, pubmed.ncbi.nlm.nih.gov/32799105/.

3. Hamouche, Salima. "COVID-19 and Employees' Mental Health: Stressors, Moderators and Agenda for Organizational Actions." *Emerald Open Research*, Emerald Publishing Ltd, 20 Apr. 2020, emeraldopenresearch.com/articles/2-15.

4. Islam, S.M Didar-Ul, et al. "Exploring COVID-19 Stress and Its Factors in Bangladesh: A Perception-Based Study." *Heliyon*, Elsevier, 10 July 2020, www.sciencedirect.com/science/ article/pii/S2405844020312433.

5. Y. Bai, L. Yao, et al. "Prevalence of Stress, Anxiety, Depression among the General Population during the COVID-19 Pandemic: a Systemic Review and Meta-Analysis." *Globalization and Health*, BioMed Central, 6 July 2020, globalizationandhealth.biomedcentral.com/articles/10.1186/s12992-020-00589-w?fbclid=Iw

6. Lakhan, Ram, et al. "Prevalence of Depression, Anxiety, and Stress during COVID-19 Pandemic." *Journal of Neurosciences in Rural Practice*, Thieme Medical and Scientific Publishers Pvt. Ltd., 11 Sept. 2020

7. Pollard, Michael. "Changes in Adult Alcohol Use and Consequences During the COVID-19 Pandemic in the US." JAMA Network Open, JAMA Network, 29 Sept. 2020, jamanetwork.com/journals/jamanetworkopen/article-abstract/2770975.

8. Rheenen, Tamsyn E. Van, et al. "Mental Health Status of Individuals with a Mood-Disorder during the COVID-19 Pandemic in Australia: Initial Results from the COLLATE Project." Journal of Affective Disorders, Elsevier, 2 July 2020, www.sciencedirect.com/science/article/pii/S0165032720323958.

9. ND; Wang Q; Xu R; Volkow. "Increased Risk of COVID-19 Infection and Mortality in People with Mental Disorders: Analysis from Electronic Health Records in the United States." *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)*, U.S. National Library of Medicine, 7 Oct.

2020, pubmed.ncbi.nlm.gov/33026219/.

10. Gonzalez-Blanco, Leticia, et al. "COVID-19 Lockdown in People with Severe Mental Disorders in Spain: Do They Have a Specific Psychological Reaction Compared with Other Mental Disorders and Healthy Controls?" *PubMed*, U.S. National Library of Medicine, 25 July 2020, pubmed.ncbi.nlm.nih.gov/32771308/.

11. Hammen, Constance, et al. "Chronic and Acute Stress and the Prediction of Major Depression in Women." *Wiley Online Library*, John Wiley & Sons, Ltd, 3 June 2009, onlinelibrary.wiley.com/doi/abs/10.1002/da.20571.

12. Bale, Tracy L. "Stress Sensitivity and the Development of Affective Disorders." Hormones and Behavior, Academic Press, 8 Aug. 2006, www.sciencedirect.com/science/article/ pii/S0018506X06001826.

13. Drevets, Wayne C. "Neuroimaging Abnormalities in the Amygdala in Mood Disorders." The New York Academy of Sciences, John Wiley & Sons, Ltd, 24 Jan. 2006, nyaspubs.onlinelibrary.wiley.com/doi/abs/10.1111/j.1749-6632.2003.tb07098.x.

14. Morris, Matthew C., et al. "Cortisol Responses to Psychological Stress Predict Depression Trajectories: Social-Evaluative Threat and Prior Depressive Episodes as Moderators." *Journal of Affective Disorders*, Elsevier, 1 Aug. 2012, www.sciencedirect.com/science/article/pii/S0165032712004429.

15. Bartolomucci, Alessandro, and Rosario Leopardi. "Stress and Depression: Preclinical Research and Clinical Implications." *P.L.O.S One*, Public Library of Science, 30 Jan. 2009, journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0004265.

16. Nikolova, Y S, et al. "Divergent Responses of the Amygdala and Ventral Striatum Predict Stress-Related Problem Drinking in Young Adults: Possible Differential Markers of Affective and Impulsive Pathways of Risk for Alcohol Use Disorder." *Molecular Psychiatry*, vol. 21, no. 3, 2015, pp. 348-365., doi:10.1038/mp.2015.85.

17. Park, Crystal L, et al. "The Daily Stress and Coping Process and Alcohol Use Among College Students." *Journal of Studies on Alcohol*, vol. 65, no. 1, 2004, pp. 126-135., doi:10.15288/jsa.2004.65.126.

18. Dixit, Anita R., et al. "Prospective Study of Depression and the Risk of Heavy Alcohol Use in Women." *American Journal of Psychiatry*, American Psychiatric Association, 1 May 2020, ajp.psychiatryonline.org/doi/full/10.1176/appi.ajp.157.5.751.

19. McHugh, R Kathryn, and Roger D Weiss. "Alcohol Use Disorder and Depressive Disorders." *National Center for Biotechnology Information*, National Institute on Alcohol Abuse and Alcoholism, 1 Jan. 2019, www.ncbi.nih.gov/pmc/articles/PMC6799954/.

20. Bolton, James M., et al. "Self-Medication of Mood Disorders with Alcohol and Drugs in the National Epidemiologic Survey on Alcohol and Related Conditions." *Journal of Affective Disorders*, Elsevier, 11 Nov. 2008, www.sciencedirect.com/science/pii/S016503270800400X.

21. Bradizza, Clara M., et al. "Difficulties in Emotion Regulation in Treatment-Seeking Alcoholics with and without Co-Occuring Mood and Anxiety Disorders." *Addictive Behaviors*, Pergamon, 29 Dec. 2017, www.sciencedirect.com/science/article/pii/S0306460317304987?via=ihub

22. Tran, Thach Duc, et al. "Alcohol Use and Mental Health Status during the First Months of COVID-19 Pandemic in Australia." *Journal of Affective Disorders*, Elsevier, 7 Sept. 2020, www.sciencedirect.com/science/article/pii/S0165032720327075.

23. Bruce, Steven E., et al. "Altered Emotional Interference Processing in the Amygdala and Insula in Women with Post-Traumatic Stress Disorder." NeuroImage: Clinical, Elsevier, 13 Nov. 2012, www.sciencedirect.com/science/article/pii/S2213158212000320.

24. H. Ackermann, A. Riecker, et al. "Contributions of the Insula to Cognition and Emotion." *Neuropsychology Review*, Springer US, 18 Jan. 2014, link.springer.com/article/10.1007/s11065-014-9246-9.

25. Gardini, Simona, and Annalena Venneri. "Reduced Grey Matter in the Posterior Insula as a Structural Vulnerability or Diathesis to Addiction." *Brain Research Bulletin*, Elsevier, 7 Dec. 2011, www.sciencedirect.com/science/article/pii/S0361923011003583.

26. Critchley, Hugo D, et al. "Neural Systems Supporting Interoceptive Awareness." Nature Publishing Group, 18 Jan. 2004, www.nature.com/articles/nn1176.

27. Herwig, Uwe, et al. "Neural Correlates of 'Pessimistic' Attitude in Depression." *Psychological Medicine*, Cambridge University Press, 7 May 2010, www.zora.uzh.ch/id/eprint/ 25834/.

28. Kandilarova, Sevdalina et al. "Altered Resting State Effective Connectivity of Anterior Insula in Depression." Frontiers, Frontiers, 28 Feb. 2018, www.frontiersin.org/articles/10.3389/ fpsyt.2018.00083/full.

29. Lippard, Elizabeth T., et al. "Brain Circuitry Associated with the Development of Substance Use in Bipolar Disorder and Preliminary Evidence for Sexual Dimorphism in Adolescents." Journal of Neuroscience Research, vol. 95, no. 1-2, 2016, pp. 777-791, doi: 10.1002/jnr.23901.

30. Spitzer, R L, J B Williams, M Gibbon and M B First (1992). "The Structured Clinical Interview for DSM-III-R (SCID). I: History, Rationale, and Description." *Arch Gen Psychiatry 49*(8): 624-629.

31. Beck, A T, et al. "Beck Depression Inventory." *PsycTests*, American Psychological Association, 1996, psychnet.apa.org/doiLanding?-doi=10.1037&2Ft00742-000.

32. Collins, R L, et al. "Social Determinants of Alcohol Consumption: the Effects of Social Interaction and Model Status on the Self-Administration of Alcohol." *Journal of Consulting and Clinical Psychology*, U.S. National Library of Medicine, Apr. 1985, pubmed.ncbi.nlm.nih.gov /3998247/.

33. Bromet, Evelyn, et al. "The CoRonavIruS Health Impact Survey (CRISIS) V0.2." 2020, Attribution License: CC-BY-4.0. creativecommons.org/Licenses/by/4.0/.

34. Desikan, Rahul S, et al. "An Automated Labeling System for Subdividing the Human Cerebral Cortex on MRI Scans into Gyral Based Regions of Interest." PubMed, U.S. National Library of Medicine, 1 July 2006, pubmed.ncbi.nlm.nih.gov/16530430/.

35. Fournier, Jay C, et al. "Amygdala and Whole-Brain Activity to Emotional Faces Distinguishes Major Depressive Disorder and Bipolar Disorder." Wiley Online Library, John Wiley & Sons, Ltd, 1 Aug. 2013, onlinelibrary.wiley.com/doi/abs/10.1111/bdi.12106.

36. Myin-Germeys, I., et al. "Emotional Reactivity to Daily Life Stress in Psychosis and Affective Disorder: an Experience Sampling Study." Wiley Online Library, John Wiley & Sons, Ltd, 17 Jan. 2003, onlinelibrary.wiley.com/doi/abs/10.1034/j.1600-0447.2003.02025.x.

TABLES AND FIGURES

	Healthy Comparison (N=18)	Mood disorder (N=19)	p-value
Demographics			
Bipolar Disorder (%)	N/A	11 (57)	N/A
Major Depressive Disorder (%)	N/A	8 (42)	N/A
Mean Age (SD)	22 (2)	23 (2)	0.38
Number of Females (%)	12 (67)	14 (74)	0.73[z]
Caucasian/White (%)	11 (61)	9 (47)	0.51[z]
Hispanic/Latino (%)	2 (11)	6 (32)	0.23[f]
African American/Black (%)	0 (0)	1 (5)	1.00[f]
Asian (%)	1 (6)	2 (11)	1.00[f]
Mixed Race (%)	4 (22)	1 (5)	0.18[f]
Mean Years Between Baseline and COVID follow-up (SD)	0.79 (0.34)	0.75 (0.35)	0.72
Alcohol/Cannabis Use Disorders			
Current Alcohol Use Disorder, severe (%)	0 (0)	2 (11)	0.49[f]
Past Alcohol Use Disorder, mild (%)	0 (0)	3 (16)	0.23[f]
Current Cannabis Use Disorder, mild (%)	2 (11)	3 (16)	1.00[f]
Past Cannabis Use Disorder, mild (%)	1 (6)	3 (16)	0.60[f]
Past Cannabis Use Disorder, moderate (%)	1 (6)	0 (0)	0.49[f]
Past Cannabis Use Disorder, severe (%)	0 (0)	1 (5)	1.00[f]
Other Clinical Factors and Comorbidities			
Anxiety Disorders[h] (%)	1 (6)	8 (42)	0.02[z]
Rapid Cycling[g] (%)	N/A	4 (21)	N/A
Lifetime Psychosis (%)	N/A	5 (26)	N/A
Lifetime suicide attempt (%)	N/A	7 (37)	N/A

Table 1: Demographic and clinical factors at baseline assessment stratified by group for behavioral analyses.

	Healthy Comparison (N=9)	Mood disorder (N=6)	p-value
Demographics			
Bipolar Disorder (%)	N/A	5 (83)	N/A
Major Depressive Disorder (%)	N/A	1 (17)	N/A
Mean Age (SD)	22 (1)	22 (1)	0.40[z]
Number of Females (%)	5 (56)	3 (50)	0.83[f]
Caucasian/White (%)	7 (78)	5 (83)	1.00[z]
Asian (%)	0 (0)	1 (17)	0.40[f]
Mixed Race (%)	2 (22)	0 (0)	0.49[f]
Mean Years Between Baseline and COVID follow-up (SD)	0.78 (0.48)	0.97 (0.05)	0.29
Alcohol/Cannabis Use Disorders			
Past Alcohol Use Disorder, mild (%)	0 (0)	1 (17)	0.40[f]
Current Cannabis Use Disorder, mild (%)	2 (22)	1 (17)	1.00[f]
Past Cannabis Use Disorder, mild (%)	1 (11)	2 (33)	0.53[f]
Other Clinical Factors and Comorbidities			
Anxiety Disorders (%)	1 (11)	1 (17)	1.00[f]
Rapid Cycling (%)	N/A	3 (50)	N/A
Lifetime Psychosis (%)	N/A	2 (33)	N/A
Lifetime suicide attempt (%)	N/A	3 (50)	N/A
Mood State on Day of Scan			
Euthymic (%)	9 (100)	4 (67)	0.14[f]
Depressed (%)	0 (0)	1 (17)	0.40[f]
Hypomanic (%)	0 (0)	1 (17)	0.40[f]

Table 2: Demographic and clinical factors stratified by group for brain imaging analyses.

In tables 1 and 2, between-group (mood disorder vs. healthy comparison) differences in age and years between baseline and COVID follow-up were compared using a two-sample t-test. All other factors were examined with Chi-Square or Fisher Exact tests, as appropriate. Between-group differences in bipolar disorder, major depressive disorder, rapid cycling, lifetime psychosis, and lifetime suicide attempt were not assessed because these factors were considered an exclusion criterion, and thus not present, in the typically developing group. [z] represents p-value calculated with a Chi-Square Test. [f] represents p-value calculated with Fisher exact test. [g] Rapid Cycling reflects past-year rapid cycling in bipolar disorder participants. [h] Anxiety disorders included generalized anxiety disorder and panic disorder.

	Healthy Comparison (N=18)	Mood disorder (N=19)	p-value
Baseline Assessment			
Beck Depression Score[b] (SD)	6 (7)	11 (8)	0.14
Total Drinks[d] (SD)	8 (12)	9 (9)	0.81
Total Drinking Days[d] (SD)	2 (2)	3 (2)	0.46
COVID Follow-Up Assessment			
Beck Depression Score[b] (SD)	8 (7)	16 (9)	0.005*
Total Drinks[d] (SD)	5 (5)	11 (18)	0.19
Total Drinking Days[d] (SD)	2 (1)	2 (3)	0.47

Table 3: Clinical mood symptoms and alcohol use during heaviest drinking week at baseline and COVID follow-up assessments. Between-group differences in Beck Depression score, Total Drinks/Heaviest Drinking Week, and Total Drinking Days/Heaviest Drinking Week were compared using two-sample t-tests. [b] Recent depressive symptoms were measured with the Beck Depression Inventory. [d] Recent alcohol use was measured with the Daily Drinking Questionnaire adapted for the heaviest week over the past 30 days (DDQ-H).

	Healthy Comparison (N=18)	Mood disorder (N=19)	p-value
COVID-Related Stress[c] (SD)	12 (4)	11 (4)	0.75
Change in Beck Depression Score[b] (SD)	5 (7)	9 (13)	0.29
Change in Total Drinks[d] (SD)	-4 (10)	1 (13)	0.22
Change in Total Drinking Days[d] (SD)	-0.28	-0.37	0.89

Table 4: COVID-related stress, change in depression pre-COVID to during-COVID, and change in alcohol use pre-COVID to during-COVID. Between-group differences in COVID-related stress, change in Beck Depression score, change in Total Drinks/Heaviest Drinking Week, and change in Total Drinking Days/Heaviest Drinking Week were compared using two-sample t-tests. [c] COVID-related stress was measured with the Coronavirus Health Impact Survey V0.2. [b] Depressive symptoms were measured with the Beck Depression Inventory. [d] Alcohol use was measured with the Daily Drinking Questionnaire adapted for the heaviest week over the past 30 days (DDQ-H).

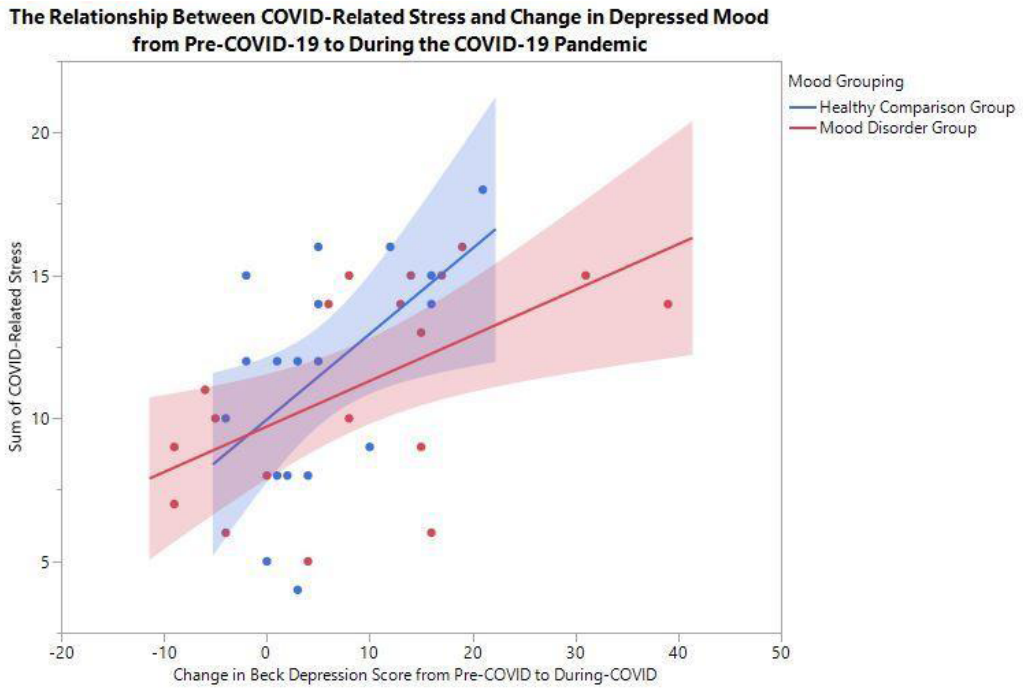


Figure 1: The relationship between COVID-related stress and change in Beck Depression Inventory score from pre-COVID-19 to during the COVID-19 pandemic. Greater total COVID-related stress is associated with greater increase in Beck Depression score across all participants (main effect of COVID-related stress, p = 0.002). There is a positive relation between COVID-related stress and change in depression within both the mood disorder group (p = 0.03) and the healthy comparison group (p = 0.04).

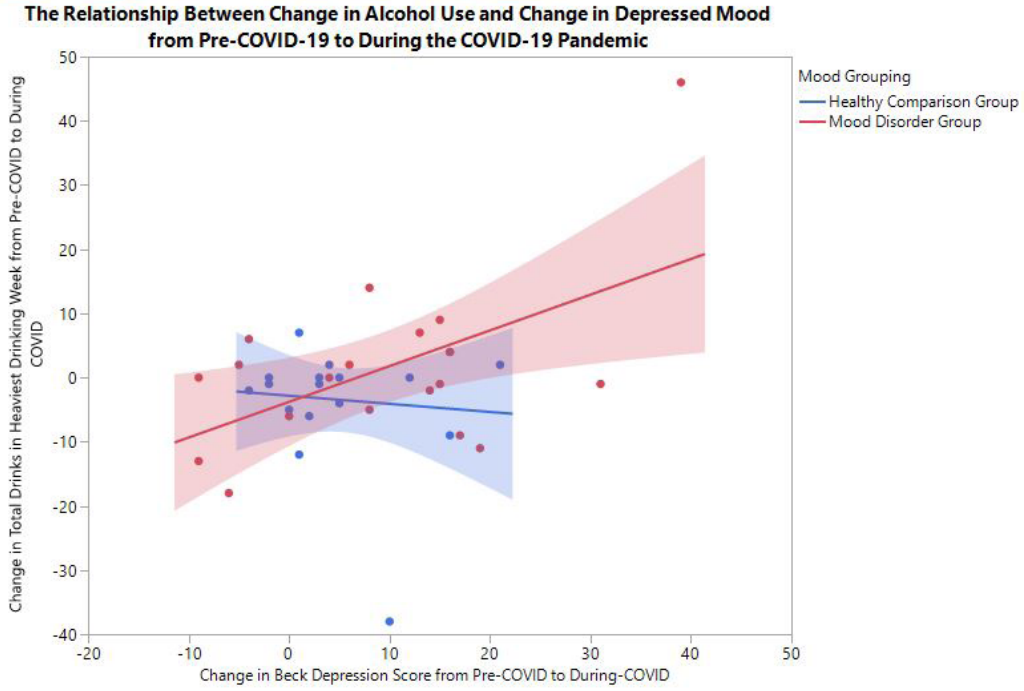


Figure 2: The relationship between change in total number of drinks consumed during heaviest drinking week from pre-COVID to during-COVID and change in Beck Depression score from pre-COVID to during COVID. Greater increase in depression is associated with greater increase in alcohol use (main effect of change in depression score, p = 0.03). There is a positive relation between change in depression and change in drinking within the mood disorder group (p = 0.04, significant), and no significant relation between change in depression and change in drinking within the healthy comparison group (p = 0.63).

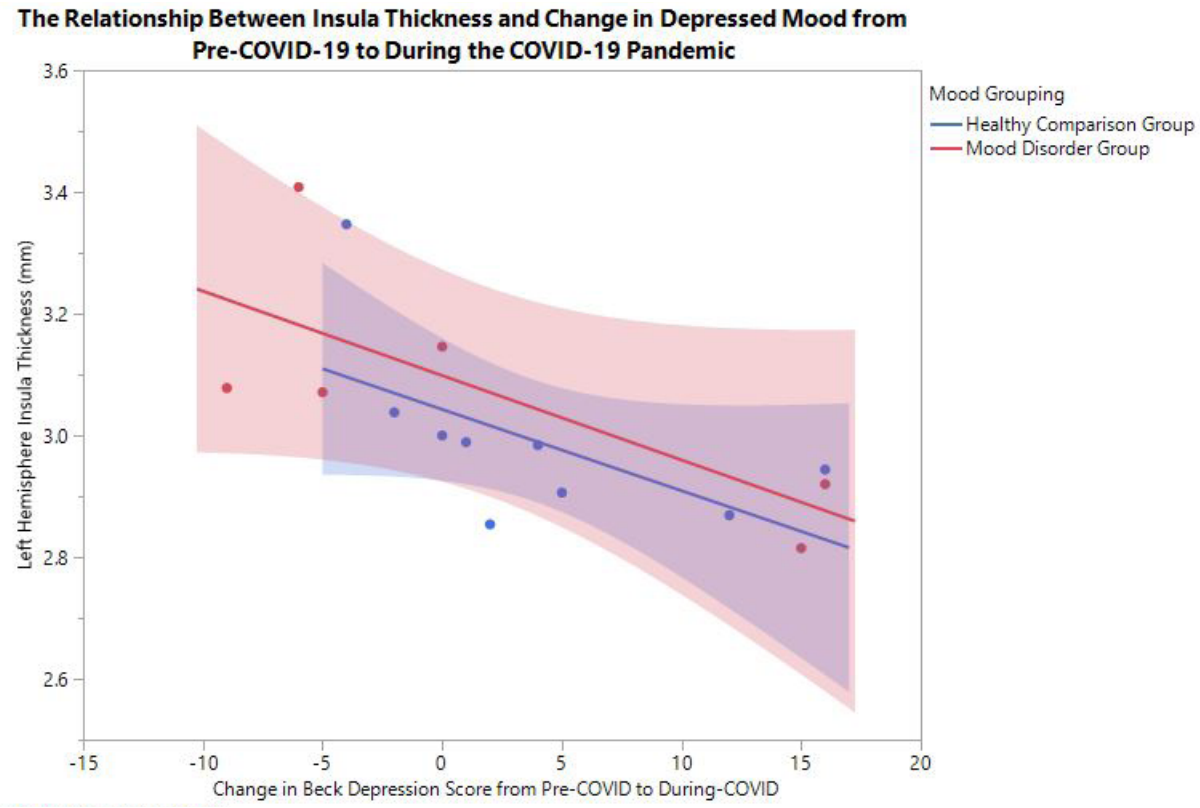


Figure 3: The relationship between left hemisphere insula thickness and change in Beck Depression Inventory score from pre-COVID to during-COVID. There is a significant main effect of left insula thickness on change in depression ($p = 0.002$) such that lower measures of left insula thickness are associated with greater increases in depression during the pandemic. There is a negative relation between insula thickness and change in depression within the healthy comparison group ($p = 0.04$, significant), and no significant relation between insula thickness and change in depression within the mood disorder group ($p = 0.09$).

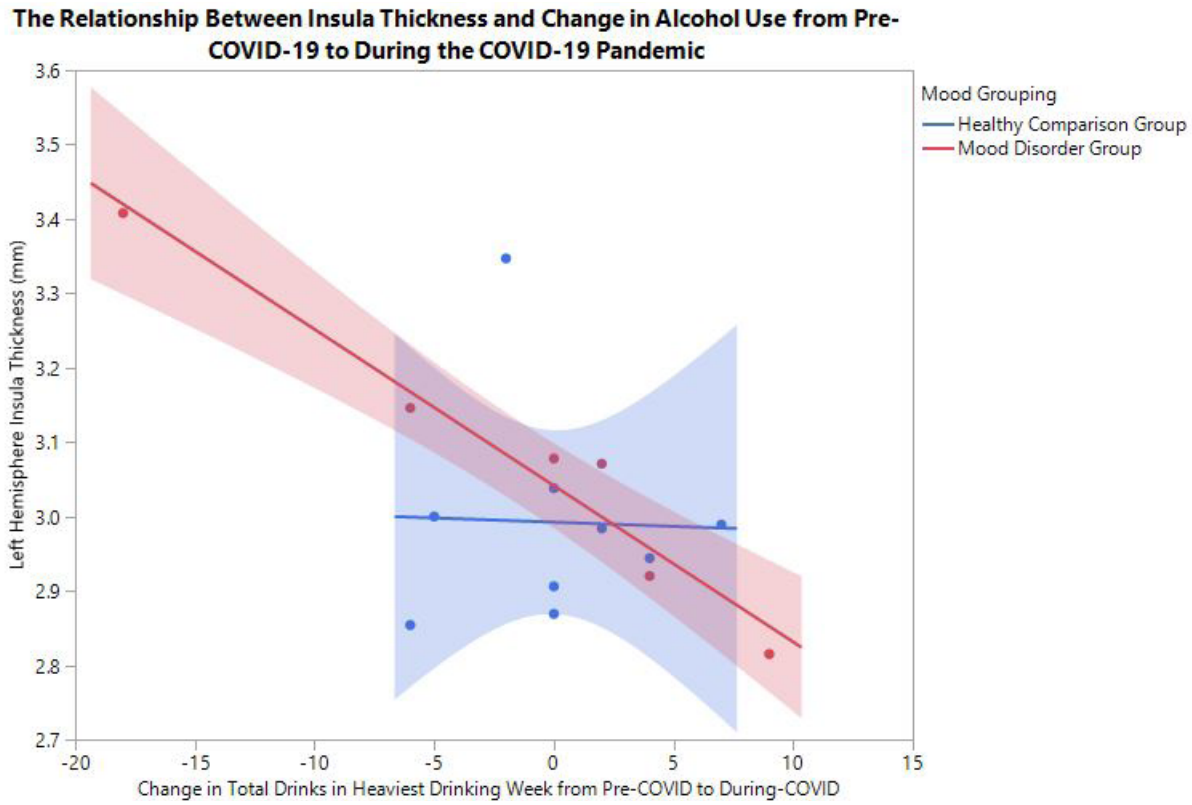


Figure 4: The relationship between left hemisphere insula thickness and change in total number of drinks consumed in heaviest drinking week from pre-COVID to during-COVID. There is a significant main effect of left insula thickness on change in total drinks consumed in heaviest drinking week ($p = 0.008$) such that lower measures of left insula thickness is associated with greater increases in alcohol use during the pandemic. There is a significant group by left insula thickness interaction on change in total number of drinks consumed in heaviest drinking week ($p = 0.0038$). There is a significant negative relation between left insula thickness and change in total number of drinks consumed in heaviest drinking week in the mood disorders group ($p = 0.002$), but no relation in the healthy control group ($p = 0.28$).