ABSTRACT

Background: Presently the diagnosis of personality disorders, particularly those that are concerned with manipulative traits such as those associated with Antisocial Personality Disorder, are extremely time consuming and variable from clinician to clinician. With the rise of social media platforms personal writings over time have become more widely available. If a new system of analysis were to be utilized as a way to help clinical assessments it could significantly reduce the time invested as well as the reliability of the data. Aim: This study was meant to investigate whether there are quantitative differences in language patterns of those identified with a personality disorder (through the use of the MCMI-III) and those without. Approach: Deidentified transcripts of the Adult Attachment Interview were formatted so that text-analysis could be run using the R-studio (Version 4.1.103) software. The transcripts that met the criteria were then analyzed using a lexical dictionary created by Saif Mohammad and Peter Turney. There were differences across conditions. However, there were differences between the two groups; they were just not considered statistically significant. A potential reason for this could be sample size. When a power analysis was run, it showed that if a total of forty-four matched transcripts could be found it would give a better potential to overcome the relatively low power of this area of study. This study showed expected results concerning the content and sentiment analysis. There were differences in content across conditions, when looking at the words most frequently used. The Bigram analysis, showed that although those in Personality Disorder Condition used more negative words both conditions used the negative words similarly, when looking at these words in pairs. Analysis that had the greatest significance were those in the sentiment-analysis. The dictionary that was used was created with a lexical dictionary that takes into account the context of the words in the interview questions in order for the results to be considered more reliable.

INTRODUCTION

The diagnosis of almost all disorders defined in the DSM-5 are based primarily on some form of intensive interview, wherein a checklist is a common assessment tool, as well as a detailed history of the client. There are some controversies on the overall reliability and validity of diagnosis from clinician to clinician for some disorders. Antisocial Personality Disorder is one that tends to draw some concern mainly due to the manipulative nature of the disorder (Sneiderman, 2006). Now in this golden age of technology, where more comprehensive personal information is available; if this data was able to be run through a new text analysis system created to ensure an accurate diagnosis would be potentially be able to be given in less time. To date it has been established that there are differences, for both verbal and written language, when distinguishing between: “psychopathic”, “hysteric”, “manic” as well as “cognitive dysfunction” (Lorenz & Cobb, 1954; Freyberg et al., 2014). This study looked to see if quantitative differences in text complexity, common themes, as well as content/sentiment analysis could be picked up in a relatively small sample of data.

METHODS

Sixty-seven deidentified transcripts, from the Adult Attachment Interview, were obtained from Professor Brent Frier (2015). All individuals that partook in the original study also took the Million Clinical Multiaxial Inventory (MCMI-III), which is an assessment that evaluates the potential that each individual has a personality disorder. Thirteen individuals high enough on the Antisocial Personality Disorder scale in order to meet the criteria for that disorder. Out of the remaining fifty-four transcripts, thirteen transcripts were randomly picked to act as a matched control group. All of these transcripts were formatted so that they could be run in R-studio (Version 1.4.103), wherein text-analysis was done. Both the Bigram and Sentiment Analysis required intensive filtering in order for them to be considered valid for this form of analysis.

RESULTS

There were differences across conditions. However, there were differences between the two groups. There were significant differences between the two groups on the Antisocial Personality Disorder scale in order to meet the criteria for that disorder. Out of the remaining fifty-four transcripts, thirteen transcripts were randomly picked to act as a matched control group. All of these transcripts were formatted so that they could be run in R-studio (Version 1.4.103), wherein text-analysis was done. Both the Bigram and Sentiment Analysis required intensive filtering in order for them to be considered valid for this form of analysis.

Table 1: Measures of Text Complexity

<table>
<thead>
<tr>
<th>Measure</th>
<th>ASPD</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>M-Word Length</td>
<td>3.680</td>
<td>3.708</td>
</tr>
<tr>
<td>Number of Words</td>
<td>90436</td>
<td>97471</td>
</tr>
<tr>
<td>Lexical Density</td>
<td>0.039</td>
<td>0.0318</td>
</tr>
<tr>
<td>Lexical Diversity</td>
<td>3544</td>
<td>3097</td>
</tr>
</tbody>
</table>

*Note: This table provides the results of the sentiment analysis done on both groups with the ASPD condition and WITHOUT control condition using the RNC lexical dictionary created by Saif Mohammad and Peter Turney. There were differences in the common themes noted in each condition. The ASPD condition incorporated words associated with anger, fear, sadness, and disgust as well as used them more often. Whereas the Control condition used the aforementioned themes less, and had more positive and joyful themes. This lexical dictionary was adjusted to have more accurate rendition of the Adult Attachment Interviews.

CONCLUSIONS

Overall, the results from this study are ambiguous. At this time there were no significant differences when looking at measures of text complexity compared across conditions. However, there were effects between the two groups; they were just not considered statistically significant. A potential reason for this could be sample size. When a power analysis was run, it showed that if a total of twenty-four matched transcripts could be found it would give a better potential to overcome the relatively low power of this area of study. This study showed expected results concerning the content and sentiment analysis. There were differences in content across conditions, when looking at the words most frequently used. The Bigram analysis, showed that although those in Personality Disorder Condition used more negative words both conditions used the negative words similarly, when looking at these words in pairs. Analysis that had the greatest significance were those in the sentiment-analysis. The dictionary that was used was created with a lexical dictionary that takes into account the context of the words in the interview questions in order for the results to be considered more reliable.

REFERENCES