2018

The validity of smartphone data and its relationship to clinical symptomatology and brain biology: an exploratory analysis

BOSTON UNIVERSITY

SCHOOL OF MEDICINE

Thesis

THE VALIDITY OF SMARTPHONE DATA AND ITS RELATIONSHIP TO CLINICAL SYMPTOMATOLOGY AND BRAIN BIOLOGY: AN EXPLORATORY ANALYSIS

by

CONOR RIZZUTO

B.A., Boston University, 2015

Submitted in partial fulfillment of the

requirements for the degree of

Master of Science

2018

Approved by

First Reader ____

Dr. Simon Levy, Ph.D.
Associate Professor of Physiology and Biophysics

Second Reader ____

Dr. John Torous, M.D.
Postdoctoral Research Fellow in Clinical Informatics
Harvard Medical School

THE VALIDITY OF SMARTPHONE DATA AND ITS RELATIONSHIP TO CLINICAL SYMPTOMATOLOGY AND BRAIN BIOLOGY: AN EXPLORATORY ANALYSIS

CONOR RIZZUTO

ABSTRACT

Background: Presently, there is very little research on the clinical validity of mental health smartphone application data, its relationship to brain biology, and its ability to inform clinical decisions. This paper seeks to explore these relationships within a sample of schizophrenic patients through the analysis of data collected on the mental health smartphone application *Biewe*.

Objectives: To validate mental health smartphone applications and support their potential to augment clinical practice.

Methods: The application involved a series of 21 questions from several questionnaires including Patient Health Questionnaire-8 (PHQ-8), Generalized Anxiety Disorder-7 (GAD-7), Warning Signals Scale (WSS), Pittsburgh Sleep Quality Index, and the psychosis subscale of the Mini Mental State Examination. Data was collected over a period of 3 months, and patients attended a total of 4 clinic visits during this timeframe. Seven study participants also had brain scan data available from the BSNIP, PARDIP and Biceps studies currently in progress at MMHC which has been used for analysis. The structural MPRAGE T1 scans were processed using Free Surfer 6 in which thickness and volume measures were extracted. All statistical analyses on the data were carried out using R statistics software.

Results: Clinic and application responses within the same week were not significantly different from each other. The application answers, however, appeared to be more sensitive to structural abnormalities in the brain. Symptoms defined as a lack of normal emotional responses (i.e. negative symptoms of schizophrenia) were negatively correlated to home time and positively correlated to distance travelled, which was a counterintuitive result.

Conclusions: The results show that mobile monitoring has the potential to be a valid and reliable method of data collection and that it may be able to augment clinical decision making.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BSNIP ....Bipolar and Schizophrenia Network for Intermediate Phenotypes

FDR……………………………………………………………..False Discovery Rate

GAD-7……………………………………………………Generalized Anxiety Disorder-7

MINI……………………………………………………Mini Mental State Examination

MMHC..Massachusetts Mental Health Center

MP RAGE...Magnetization-Prepared Rapid Gradient-Echo

PARDIP………Psychosis and Affective Research Domains and Intermediate Phenotypes

PHQ-8…………………………………………………...Patient Health Questionnaire-8

WSS…………………………………………………………...Warning Signals Scale

INTRODUCTION

To date, there exists limited literature on the use of smartphone applications to assess and monitor mental health symptoms. With global smartphone application ownership cited at 25% worldwide by 2015, and 65% in the US, mobile monitoring of mental health symptoms has the potential to become a powerful tool in the clinical toolbox (Firth et al, 2015). Clinic interview scales are currently the gold standard method of assessing symptoms but cross sectional data presents several limitations, such as recall bias and interrater reliability. Mobile monitoring could be a solution to these issues by painting a longitudinal picture of a disease state, but it needs to be shown that this data is valid, and that the method of collection is safe and feasible (Palmier-Claus et al, 2012).

Of the studies that have employed the use of smartphone applications to collect data on symptoms, many have reported high satisfaction with this medium. A study involving patients with head and neck cancer demonstrated that substantial symptom reporting is possible using mobile device technology (Falchook et al, 2016), a conclusion gleaned from the impressive participation in the study (median report compliance of ~71%). Additionally, there has been investigation into the utility of smartphone applications within various branches of medicine. A study in 2016 involving an application called *Pain Buddy* was aimed at pain assessment and management in children with cancer. Several key features included daily symptom diaries, cognitive and behavioral skills training and interactive three-dimensional avatars that guided children through the program (Fortier et al, 2016). Another study discussed the use of a smart phone application to help patients manage their chronic kidney disease, and in an exit interview patients reported they felt more confident and in control of their condition (Ong et al, 2016).

The branch of medicine that can also benefit from this kind of innovation is psychiatry. Mental illness puts a huge burden on the individual and the economy, and yet there is significant unmet need in the population. It is cited that only about 25% of persons with mental health problems currently receive adequate professional help (Beiwinkel et al, 2017). It is therefore imperative that steps are taken to improve this statistic. This study recruited patients with a diagnosis of schizophrenia. Schizophrenia is a mental illness occurring in about 1% of the population and is characterized by psychotic symptoms, cognitive impairment and functional decline. While a plethora of research has been done on schizophrenia, much of the disease state including its etiology remains unknown (Feigenson et al, 2014). The generally agreed upon central dichotomy of symptoms present in patients with schizophrenia is the distinction between positive and negative symptoms. Positive symptoms are defined as hallucinations, delusions, disorganized thoughts and behavior and related aggression (Carbon et al, 2014). Negative symptoms are defined as a lack of normal emotional responses, thought processes and behaviors and include flat affect, apathy, anhedonia and poor concentration. These symptoms contribute to poor long term outcomes through a reduction of quality of life and increased burden of illness (Chue et al, 2014). It has been strongly suggested that the anterior cingulate cortex is implicated in the pathogenesis of schizophrenia, and more specifically, with the negative symptoms of schizophrenia. Negative symptom comparisons in this study were therefore focused on this region. Studies have also shown correlations between positive symptoms and gray matter volume abnormalities in the caudate, corpus callosum, fusiform gyrus and to overall gray matter. These volumes were the focus of the positive symptom comparisons.

Schizophrenia is a serious mental illness that is often hard to find adequate care for. Symptoms usually present in early adulthood or late adolescence and males tend to have a worse prognosis, including more negative symptoms and less chance of a full recovery. A systematic review of studies on schizophrenia has determined that patients with schizophrenia have smaller whole brain volumes, with the greatest impact on grey matter in the frontal and temporal lobes. Additionally, 80% of patients recover after their first episode, but less than 20% of those patients will never have a recurrent episode (Picchioni et al, 2007). Furthermore, a study by Emsley et al (2013) stated that multiple relapses characterize the course of illness for most patients living with schizophrenia, but the nature of these relapses has not be extensively researched. In addition to general assessment and management of symptoms, it is the hope that mobile data will eventually help us learn when patients are at their highest risk of a relapse.

SPECIFIC AIMS

In a previous study on patients with schizophrenia where smartphone and in person clinic responses were compared, 5 items showed moderate to strong associations to corresponding items from interview scales. These items included delusions, hallucinations, suspiciousness, anxiety and hopelessness (Palmier-Claus et al, 2012). A similar goal was sought to be achieved in this study.

The primary goals of the study were to assess overall engagement with a symptom reporting application, to determine whether responses were consistent across application and clinic (i.e. there was no evidence of a tendency for under or over reporting), to introduce passive measures as a useful future digital biomarker, to evaluate the sensitivity of the application and clinic scales to structural abnormalities in the brain and finally, to provide a case for the integration of mobile monitoring into the world of mental healthcare.

METHODS

Data Collection

The sample included schizophrenia patients (n = 19), 15 of whom participated in the smartphone side of the study. The application involved a series of 21 questions from several questionnaires including PHQ-8, GAD-7, Warning Signals Scale (WSS), Pittsburgh Sleep Quality Index, and the psychosis subscale of the MINI. The scales on the former four questionnaires were 0-3, while the psychosis subscale is in the format of "Yes" or "No" answers. These were recorded as "0" for "No" and "1" for yes. The application prompted participants to answer 7 questions a day every Monday, Wednesday and Friday of each week so that at the end of each week they had answered all 21 questions. All participants were given the option to postpone surveys for up to two days after they were first presented. This survey data is referred to throughout the text as "active data" to make the distinction between survey answers and "passive data". Passive data was collected on the *Biewe* application through smartphone GPS data and was collected during the entirety of the 3-month study as long as the patient's phone was on. The passive measures of interest were home time, significant locations visited, distance travelled, outgoing and incoming texts (Table 1). Participants were asked to attend a total of 4 clinic visits spaced 1 month apart during this timeframe. Each patient's first clinic visit acted as a baseline where they were taught how to set up the application and how to use it. At each visit, they were also asked to complete the PHQ -8, GAD7, Warning Signals Scale (WSS), Pittsburgh Sleep Quality Index, and the psychosis subscale of the

MINI. Seven study participants also had brain scan data available from the BSNIP, PARDIP and Biceps studies currently in progress at MMHC which was used for analyses. The structural MP RAGE T1 scans were processed using Free Surfer 6, an open source software suite for processing and analyzing brain MRI images, in which thickness and volume measures were extracted.

Table 1. Passive Measures. Descriptions of the passive measures of interest in this study, derived from the work done by Canzian et al (2015).

Passive Measure	***Description***	***Method of Calculation***
Distance Travelled	Total distance covered each day over the course of the study, where d(C_i + C_i + 1) is the geodesic distance between the latitude-longitude pairs C_i and C_i+1. Units are arbitrary.	$N_{sig}(t_1, t_2) = \sum_{i=1}^{N(t_1,t_2)} d(C_i, C_{i+1})$
Home Time	"Home" is defined as the cluster in which the user is found most often at 02:00, 06:00 and 20:30 on weekdays. Therefore, home time is calculated as the time intervals in which the participant is present at the coordinates C_H, of this cluster.	$H_t = \sum t(C_H)$ given in minutes/day
Significant Locations Visited	"Significant Location" is assigned to the 10 most visited places among all places visited by the user in the time period of the study.	$N_{sig}(t_1, t_2) = \sum_{j=1}^{10} min\left\{ \sum_{i=1}^{N(t1,t2)} 1_{isj}, 1 \right\}$
Outgoing Texts	Total texts sent each day over the course of the study.	Direct data retrieval from database. Apple phones do not allow this.
Incoming Texts	Total texts received each day over the course of the study.	Direct data retrieval from database. Apple phones do not allow this.

Survey Participation

To approximate total participation, the total time each participant was present in the study was used to determine the maximum number of surveys we know were presented to the patient, and the actual number of surveys taken was divided by this number. Each patient then had a survey participation rate percentage that was used for a correlation analysis. The main passive metric that was of interest here was home time in minutes, with a hypothesis that patients would be more active on their phones the more they stayed home. The participation rates were then split into two groups based on phone type (Android *vs.* Apple), and total participation rates for each were determined.

Fisher's Exact Tests

Fisher's Exact was used over a chi square analysis due to the small sample size and the resulting small number of observations for each rating (i.e. some observations were 0). Eight out of the 21 questions available on the application that were the most comparable (similar to identical) to the clinic questions were chosen as a basis for comparison to determine whether there was a significant difference in symptom reporting on the application *vs.* the gold standard clinical scales. There were 19 patients in the study, and only 15 had active data recorded on the application. One of those 15 patients was removed due to a lack of responses in close proximity (+/- 3 days) to the clinic dates. The tests included responses from a total of 47 clinic visits and 69 application answers within the same week as the clinic visits (+/- 3 days from each clinic visit). Because the study group included both Apple and Android software, phone type was used as a covariate in the Fisher analysis by separating the analysis into two groups. Gender was not considered due to the very uneven distribution in this study. The tests were carried out using the function “fisher.test” in R Studio.

Repeated Measures Correlations

Repeated measures correlations are used to determine the common within-individual association for paired measures assessed on two or more occasions for multiple individuals (Bakdash et al, 2017). The passive data totaled over 1700 entries, and because each measure was assessed on multiple occasions for each patient, it was decided that this was an appropriate statistical technique to use. For this analysis, passive measures home time, distance travelled, significant locations visited, outgoing texts and incoming texts were looked at. Only data from Android users was used due to the data collection restrictions that Apple has enforced. To carry out the analysis, the package “rmcorr” was installed into base R Studio. Results were adjusted for multiple comparisons with the false discovery rate, which is a method of controlling for type 1 errors, or false positives, when hypothesis testing with multiple comparisons.

Correlation Coefficient

Pearson's correlations were performed in the analysis of the *Biewe* data set through the package "Hmisc" with the function "rcorr". Upon examining the histograms of the passive data used in these correlations, distance travelled appeared logarithmic, thus a log transformation was performed with R Studio prior to running the correlations to make the data more uniform and to aid in any linear associations. The data was then adjusted with the false discovery rate. No covariates were included due to the extremely small sample size, but the possibilities included age, race, scanner type and phone type. The effect of gender would have likely been negligible because only one female used the smartphone application.

Symptom Scores

Positive and negative symptom scores were determined for each patient with brain scans by grouping several symptom responses and averaging these responses over the course of the study to account for short term fluctuations. The negative symptom score on the application included the responses from the following symptoms: little interest or pleasure in things, trouble concentrating and withdrawing from social interaction. The negative symptom score for the clinic included a lack of enthusiasm in doing things, trouble concentrating, unable to concentrate, and lack of interest in things. The positive symptom score for the application was a combination of difficulty thinking clearly, feeling confused or puzzled, feeling suspicious and hearing voices or seeing things. The symptoms included in the clinic positive symptom score were feeling confused, feeling paranoid now, external locus of control now, special messages now, strange beliefs now, hearing voices now and having visions now. The scores were then correlated with the volumes of brain structures that have been reported in the literature to be involved in schizophrenia. Of note, the psychosis subscale of the MINI includes “Now” and “Ever” versions of these questions. Only the “Now” versions were considered useful in this analysis.

Plots and Tables

The stacked column graph was generated with the R Studio package “ggplot2” and with the function “ggplot”. The eight symptoms were identical to the ones used in the Fisher’s exact tests. Most of the tables were generated with the package “FlexTables” in R Studio with the function “vanilla.table”. The demographic table was created in Excel.

RESULTS

Overall, the study recruited 19 participants with a diagnosis of schizophrenia. Only 15 of those patients participated in the application side of the study and of those 15, 7 had brain scans available from parallel studies. There were only 2 females in the study, the rest identifying as male. The two phone types represented in the study were Apple and Android. The two types of scanners used for Magnetic Resonance Imaging were MR750 and Siemens. This information is summarized in Table 2.

A longitudinal display of the study was created with stacked columns to capture the entirety of the study by week, and to show the contrast between one patient who was very active in reporting their symptoms *versus* one patient who was not. The stacked column display in Figure 1 is color coded by symptom and the x axis is in a week – year format, with clinic visits labeled.

Table 2. Demographics of the Study. This table of characteristics shows the number of participants that belong to each category and the percent of the total participants that this category represents. The scanner section only applies to the 7 patients with brain scan data available.

Characteristic	***Number***	***Percent***
Age		
20-29	14	74%
30-39	4	21%
40-43	1	5%
Gender		
Female	2	12%
Male	17	88%
Race		
Caucasian	8	42%
African American	3	16%
Hispanic	4	21%
East African	2	11%
East Asian	1	5%
Caribbean	1	5%
Phone Type		
Apple	8	42%
Android	11	58%
Scanner		
MR750	4	57%
Siemens	3	43%

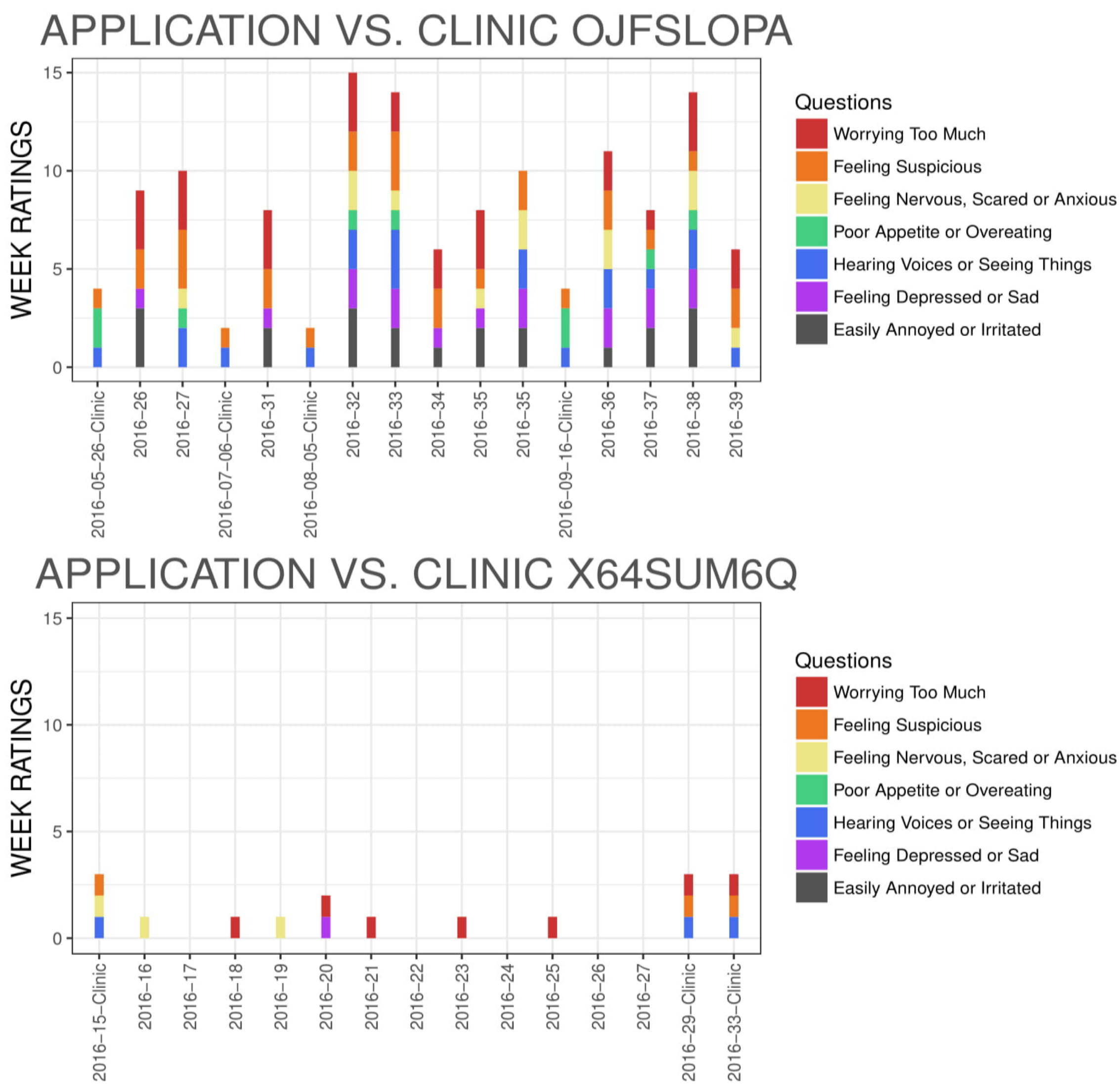


Figure 1. A Longitudinal Display of the Study. This figure is a longitudinal stacked column display of the study with the responses of two patients to seven different symptoms over the course of the study. The title of each plot shows the subject ID's of both patients to preserve anonymity. The height of each colored bar corresponds to the severity of the symptom during the week the survey was taken, on a scale of 0-3.

A total survey completion rate of 57% was determined for Android users and 49% for Apple users. Regressing survey participation rate and the passive measure of home time in minutes yielded a positive correlation with an R^2 of 0.22. This is shown in Figure

2 below.

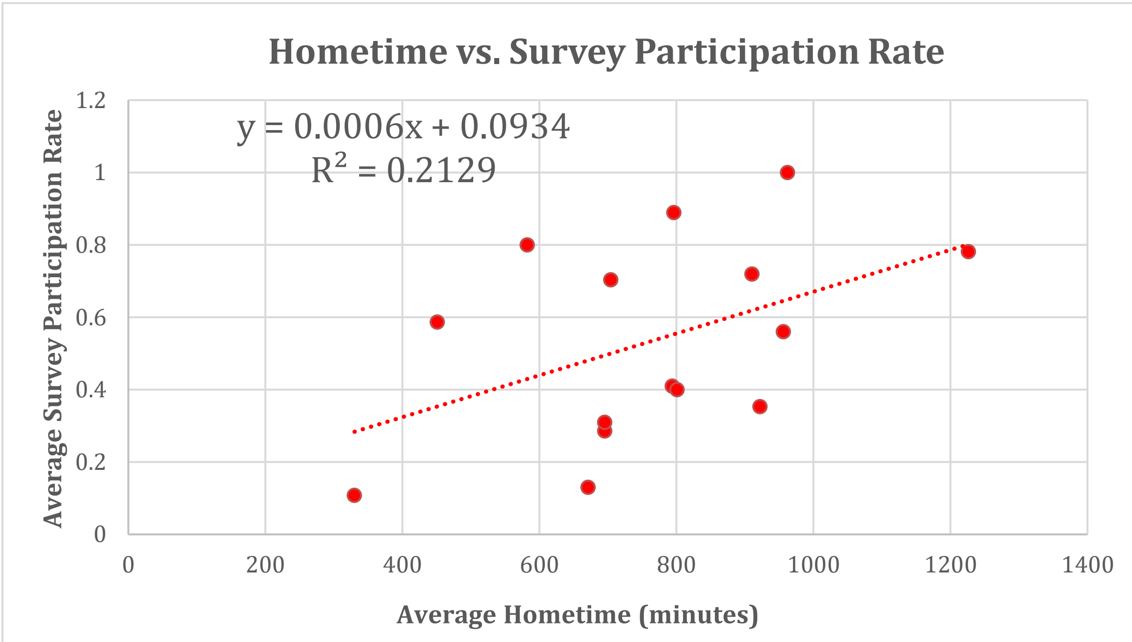


Figure 2: Home Time *vs.* Survey Participation Rate. A scatter plot using the average participation rates for the application *vs.* average home time measured in minutes.

To determine if there were any significant differences in the way symptoms were reported on the application *vs.* the clinic visits, Fisher's exact tests were performed on the frequencies of 8 of the 21 symptoms. The application responses used were within +/- 3 days of the clinic visit responses. Responses were measured with and without controlling for phone type. There were no significant findings in this analysis. The findings are separated into three tables below. Table 3 includes both phone types, Table 4 shows only Apple users, and Table 5 only Android users.

Table 3. Total Fisher's Exact Analysis. The Fisher's exact p values generated after comparing the responses to 8 different symptoms (on a scale of 0 to 3, 3 being the most severe) in clinic and on the application within +/- 3 days of each other. The times each question was answered in clinic is constant for each symptom because each visit required the patient to answer all questions, but the times answered on the application vary due to skipped days and postponed surveys.

Symptom	Times Answered in Clinic	Times Answered on Application	Fisher's Exact
Worrying Too Much	47	31	0.704
Trouble Relaxing	47	23	0.885
Feeling Nervous, Scared or Anxious	47	23	0.736
Poor Appetite / Overeating	47	23	0.361
Little Interest or Pleasure in Things	47	31	0.443
Feeling Depressed or Sad	47	26	0.301
Irritable	47	26	0.780
Feeling Bad or Guilty About Yourself	47	23	0.919

Table 4. Fisher's Exact Analysis for Apple Users. The Fisher's exact p values generated after comparing the responses to 8 different symptoms in clinic and on the *Biewe* application for Apple users.

Symptom	Times Answered in Clinic	Times Answered on Application	Fisher's Exact
Worrying Too Much	25	16	0.489
Trouble Relaxing	25	12	0.209
Feeling Nervous, Scared or Anxious	25	12	1.000
Poor Appetite / Overeating	25	12	0.290
Little Interest or Pleasure in Things	25	16	0.847
Feeling Depressed or Sad	25	14	0.348
Irritable	25	14	0.788
Feeling Bad or Guilty About Yourself	25	12	0.346

Table 5. Fisher's Exact Analysis for Android Users. The Fisher's exact p values on the response to eight different symptoms on the *Biewe* application and in the clinic for Android users.

Symptom	Times Answered in Clinic	Times Answered on Application	Fisher's Exact
Worrying Too Much	22	15	0.414
Trouble Relaxing	22	11	0.433
Feeling Nervous, Scared or Anxious	22	11	0.192
Poor Appetite / Overeating	22	11	0.864
Little Interest or Pleasure in Things	22	15	0.147
Feeling Depressed or Sad	22	12	0.682
Irritable	22	12	0.486
Feeling Bad or Guilty About Yourself	22	11	0.570

Home time *vs.* distance travelled acted to confirm that the data sensor was accurate, with an R^2 of -0.25 and an adjusted p value of <0.001. There was also a strong positive correlation between incoming and outgoing texts, with an R of 0.94 and adjusted p value of <0.001 (Table 6).

Table 6. Repeated Measures Correlation Data: The "within subject" correlation data for five passive measures: significant locations visited, home time, distance travelled, outgoing and incoming texts, with false discovery rate adjusted p values.

comparisons	r	p	p_adj
significant locations visited - hometime	-0.08	0.02	0.07
distance travelled - hometime	-0.25	8.62 x 10^-14	4.31 x 10^-13
outgoing texts - hometime	-0.03	0.346	0.576
incoming texts - hometime	-0.06	0.089	0.178
distance travelled - significant locations visited	0.07	0.037	0.093
outgoing texts - significant locations visited	0.01	0.816	0.906
incoming texts - significant locations visited	0.01	0.744	0.906
outgoing texts - distance travelled	0.00	0.907	0.907
incoming texts - distance travelled	0.03	0.435	0.621
incoming texts - outgoing texts	0.94	0	0

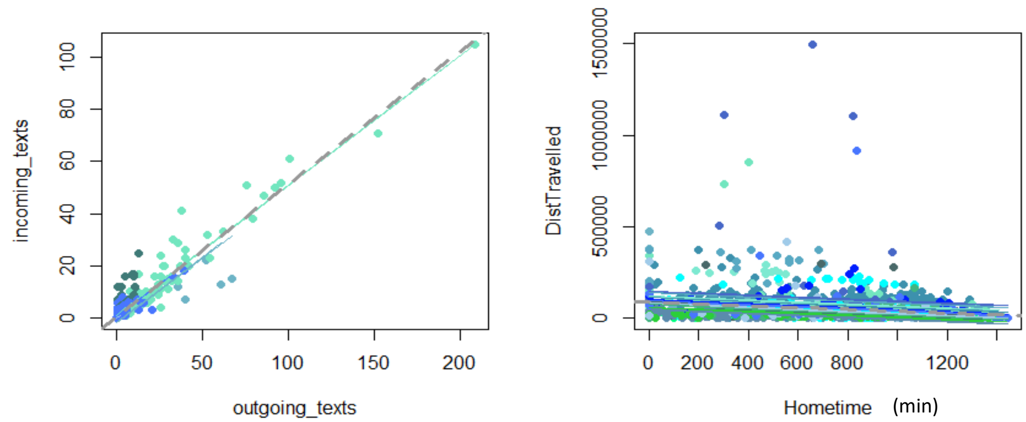


Figure 3: Repeated Measures Correlations. The **"**within subject" correlations in the passive data for incoming *vs.* outgoing texts, and home time *vs.* distance travelled. Repeated measures correlations shows every data point collected over the course of the study for all the Android users in the study.

Negative and positive symptoms were correlated against two passive measures, home time and distance travelled. Application negative symptoms scores were negatively correlated with the home time measure ($p = 0.015$, $r = -0.85$). Clinical negative symptom scores *vs.* home time, and application negative symptom scores *vs.* distance travelled, were found to be approaching significance with p values of 0.059 and 0.058, respectively.

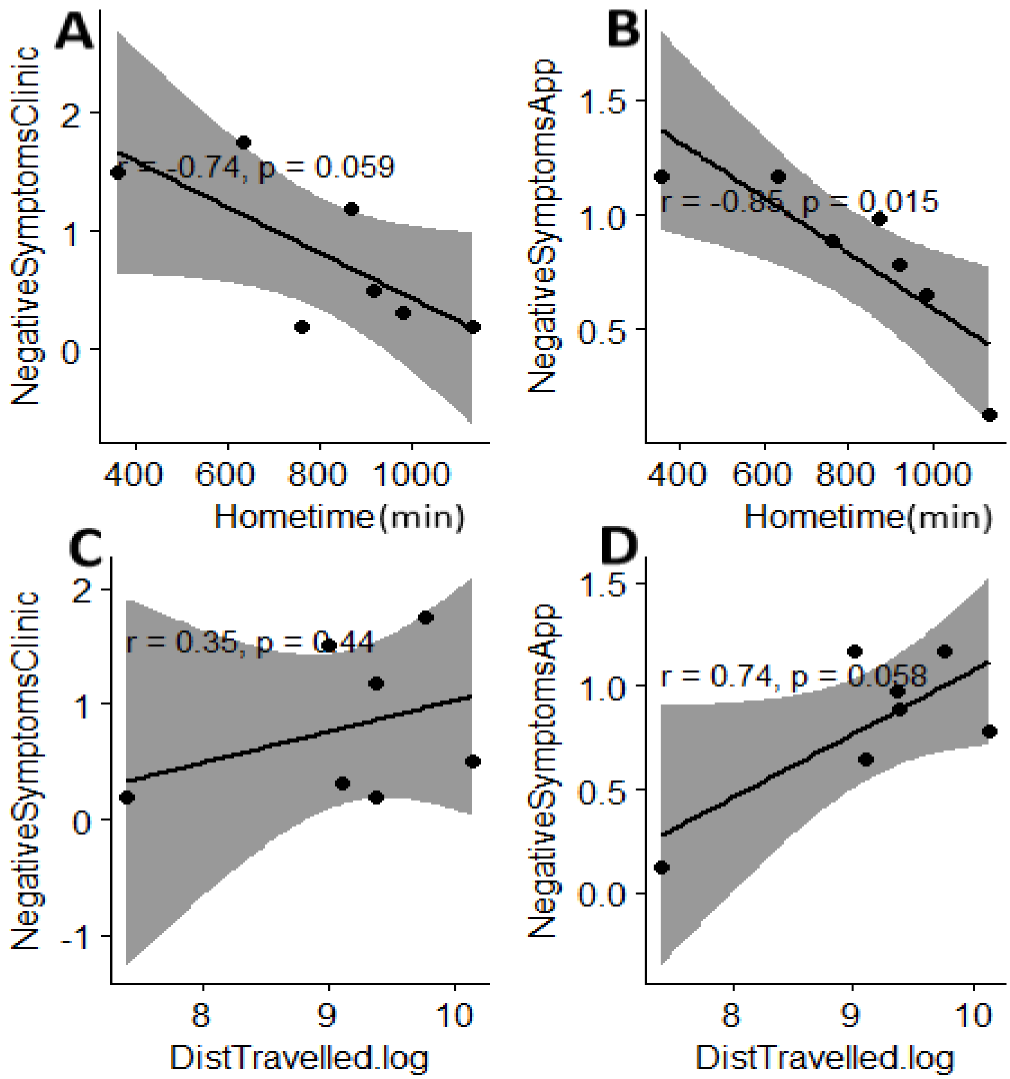


Figure 4: Scatter Plot Pearson Correlations. Four scatter plots showing the general trend agreement between the passive data measures and negative symptom scores for both the clinic and application. Plots A and B are graphs of the clinic and application negative symptom scores *versus* hometime in minutes per day. Plots C and D are the clinic and application negative symptom scores *versus* distance travelled. Distance travelled has arbitrary units and the scale in these plots is logarithmic due to the log transformation.

Table 7 shows the results of all correlations done between the passive data and symptom scores. There were no significant results found for the positive symptom scores.

Table 7. Symptom Score *vs.* Passive Data Correlations. Pearson correlations between passive measures home time and distance travelled, and positive and negative symptom scores in both the clinic and on the application. The table includes the r and p values from the analysis.

comparisons	r	p
distance travelled - clinic positive symptoms	0.00	0.997
distance travelled - app positive symptoms	0.26	0.567
distance travelled - app negative symptoms	0.74	0.057
distance travelled - clinic negative symptoms	0.35	0.435
hometime - clinic positive symptoms	-0.21	0.659
hometime - app positive symptoms	-0.32	0.490
hometime - app negative symptoms	-0.85	0.015
hometime - clinic negative symptoms	-0.74	0.058

Negative symptom scores for clinic and application were also compared to bilateral anterior cingulate volumes. The application negative symptom scores were negatively correlated with right anterior cingulate volume (r = - 0.87, p = .036) and right rostral anterior cingulate volume (r = -0.84, p = .036). The clinic negative symptom scores were negatively correlated with the right caudal anterior cingulate volume (r = -0.87, p = .04). These results are shown in Table 8 and 9.

Table 8. Application Negative Symptom Score Correlations. Pearson correlations between application negative symptom scores and various brain structures as implicated in the literature on schizophrenia. The table includes the r and p values, as well as the FDR corrected p value which adjusts for multiple comparisons.

comparisons to negative symptoms	r	p	p_adj
Left Caudal Anterior Cingulate Volume	-0.40	0.380	0.506
Left Rostral Anterior Cingulate Volume	-0.01	0.980	0.980
Right Caudal Anterior Cingulate Volume	-0.88	0.010	0.036
Right Rostral Anterior Cingulate Volume	-0.84	0.018	0.036

Table 9. Clinic Negative Symptom Score Correlations. Pearson correlations between clinic negative symptom scores and various brain structures as implicated in the literature on schizophrenia. The table includes the r and p values, as well as the FDR corrected p value which adjusts for multiple comparisons.

comparisons to negative symptoms	r	p	p _adj
Left Caudal Anterior Cingulate Volume	-0.07	0.878	0.878
Left Rostral Anterior Cingulate Volume	0.48	0.280	0.373
Right Caudal Anterior Cingulate Volume	-0.87	0.010	0.042
Right Rostral Anterior Cingulate Volume	-0.52	0.227	0.373

The positive symptom scores were correlated against a different set of brain volumes based on published literature. The application positive symptom scores were positively correlated with right fusiform volume ($r = 0.88$, $p = 0.036$), left caudate volume ($r = 0.84$, $p = 0.036$) and corpus callosum mid anterior volume ($r = 0.84$, $p =$

0.036), with sub cortical gray volume found to be trending significant post false discovery rate correction. The clinic positive symptom scores were positively correlated with right fusiform volume (r = 0.82, p = 0.046), left caudate volume (r = 0.87, p = 0.046) and corpus callosum mid anterior volume (r = 0.83, p = 0.046). These results are summarized in Tables 10 and 11.

Table 10. Application Positive Symptom Score Correlations. Pearson correlations between application positive symptom scores and various brain structures as implicated in the literature on schizophrenia. The table includes the r and p values, as well as the FDR corrected p value which adjusts for multiple comparisons.

comparisons to positive symptoms	r	p	p_adj
Left Fusiform Volume	0.26	0.573	0.573
Right Fusiform Volume	0.88	0.010	0.036
Left Caudate Volume	0.84	0.018	0.036
Right Caudate Volume	0.60	0.153	0.184
CC Mid Anterior Volume	0.84	0.018	0.036
Sub Cortical Gray Volume	0.79	0.035	0.052

Table 11. Clinic Positive Symptom Score Correlations. Pearson correlations between clinic positive symptom scores and various brain structures as implicated in the literature on schizophrenia. The table includes the r and p values, as well as the FDR corrected p value which adjusts for multiple comparisons.

comparison to positive symptoms	r	p	p_adj
Left Fusiform Volume	0.40	0.378	0.378
Right Fusiform Volume	0.82	0.023	0.046
Left Caudate Volume	0.87	0.011	0.046
Right Caudate Volume	0.68	0.094	0.113
CC Mid Anterior Volume	0.83	0.020	0.046
Sub Cortical Gray Volume	0.72	0.071	0.106

DISCUSSION

The Fisher's exact test results confirmed the hypothesis that there would be no discrepancies in symptom reporting on the application *versus* the clinic and that the responses in both environments would generally agree with each other. This was not performed on every single symptom in the study because the question sets were not identical, but it did account for symptoms across several different diagnoses including psychosis, depression and anxiety.

The repeated measures correlation to determine within-subject correlations in the passive data produced two significant results. Home time *vs.* distance travelled was expected to produce a negative correlation as long as the passive data sensor was functioning correctly, so the result of the analysis ($r = -0.25$, $p < 0.001$, Table 5) was reassuring. The strong positive correlation between incoming and outgoing texts ($r = 0.94$, $p < 0.001$, Table 5) was an interesting finding because it may suggest, at least when it comes to the Android users in this study, that these patients are highly responsive to mobile communication. This may have implications for the utility of mobile physician – patient communication in the future, but firm conclusions are hard to draw from a small sample size.

Correlating symptom scores with the passive measures of home time and distance travelled yielded counterintuitive results. The negative symptom score *versus* home time correlations for both the application and clinic appeared to agree that patients who spent more time at home had less negative symptoms (Figure 3).

We hypothesized that the more withdrawn patients would be spending more time at home, so this result was unexpected. It could be, however, that the patients with more severe disease states do not have a stable home or home life and that a stable home life and social interaction at home leads to better outcomes, which is a testament to the work done by Arsova et al (2016) in which better personal and social functioning was observed in patients who had family support. This relationship would need to be pursued with a large sample size to make any definitive conclusions.

In an effort to determine whether the application was more sensitive to picking up abnormalities in the anterior cingulate cortex, the gray matter volumes of the left and right caudal and rostral anterior cingulate were run against negative symptom scores. The application scores showed significant negative correlations to the right caudal anterior cingulate ($r = -0.87$, $p = 0.03$) and the right rostral anterior cingulate ($r = -0.84$, $p = 0.03$), while the clinic scores showed a significant negative correlation only to the right caudal anterior cingulate ($r = -0.87$, $p = 0.04$). The review study by Bersani et al cited 9 studies that evidenced a link between negative symptoms and hypo-activity of the anterior cingulate cortex, and 7 studies that did not (Bersani et al, 2018). Furthermore, a study on major depressive disorder found a significant correlation between depressive symptoms and the rostral anterior cingulate cortex (Yoshimura et al, 2010). The right sided localization could be related to a study done in 2013 that showed that when compared to controls, the right anterior cingulate gyrus gray matter volume was significantly reduced

(Takayanagi et al, 2013), though the correlation here was done without a control group. A study done in 2005 that found a negative correlation between caudal anterior cingulate cortex gray matter volume and positive symptoms (Choi et al, 2005) may disagree with the data found here even though the relationship between positive symptoms and the ACC was not explored in this study.

Positive symptom scores for the application were positively correlated with right fusiform gyrus gray matter volume ($r = 0.87$, $p = 0.03$), left caudate gray matter volume ($r = 0.84$, $p = 0.03$) and corpus callosum mid anterior gray matter volume ($r = 0.84$, $p = 0.03$). Sub cortical gray matter volume was trending after FDR correction ($r = 0.79$, $p = 0.05$). The clinic scores yielded similar results, though the sub cortical gray matter volume was not significant. A study in 2001 found that the degree of reduction in left caudate volume in schizophrenic patients after treatment with clozapine was significantly related to the extent of improvement in positive and general symptoms but not negative symptoms. The study also cited reports that typical antipsychotics caused an increase in left caudate volume, which may be an ineffective brain adaptation since large caudate volumes have been associated with more severe symptoms (Scheepers et al, 2001). Though the results could be explained by typical antipsychotic use, medication data was not available at the time of this study. Also of note is that when compared to controls, the volume of the caudate has been found to be significantly reduced (Ebdrup et al, 2010). Xiao et al (2013) discovered that patients with schizophrenia and their unaffected relatives both showed increased right fusiform gray matter volumes when compared to healthy controls, which they suggested may be evidence of subtle genetic anatomical brain deficits. The study also noted that patients with schizophrenia and their unaffected relatives have shown similar functional deficits in this area. The fusiform gyrus is implicated in face-processing tasks, so the authors proposed that the increased volumes seen in both groups could be a compensatory phenomenon in response to reduced language-related lateralization. This compensatory mechanism may be more pronounced in patients with more severe disease states as defined by the presence of more positive symptoms. As for the corpus callosum, a 1988 study showed that the mean size of the anterior region of the corpus callosum was significantly greater in schizophrenics than in controls (Uematsu et al, 1988).

The results in this study have many implications for the future of clinical practice and mobile monitoring. We showed that the application and the clinic data generally agree with each other, and in some instances the application could potentially be more sensitive than clinical scales to brain structure alterations commonly seen in schizophrenia. This study also introduces passive or functional data which can provide objective insight into human behavior and it is clear from the results shown here that passive data like home time and distance travelled have the potential to be useful digital biomarkers. Longitudinal monitoring is also free from the recall bias and interrater reliability issues introduced with cross sectional clinic data, which may make it more reliable. If mobile monitoring is introduced in the future to function in tandem with clinic visits, this could lead to fewer in person visits which would reduce the financial burden on the patient, and increase access to care. The amount of data that this kind of monitoring system has the potential to produce, however, should be considered from an ethics standpoint. The question is whether we really need this much data and it still remains to be seen whether applications like these, when used in concert with clinical practice, will lead to healthier patients at a lower cost. The data would have to be integrated into electronic medical records and it is unclear how much this would cost on a large scale, and if insurance would pay for the application, the data and the integration of the data. Patients would also have to feel comfortable with and understand the idea behind long term monitoring, which presents another potential future challenge due to privacy concerns.

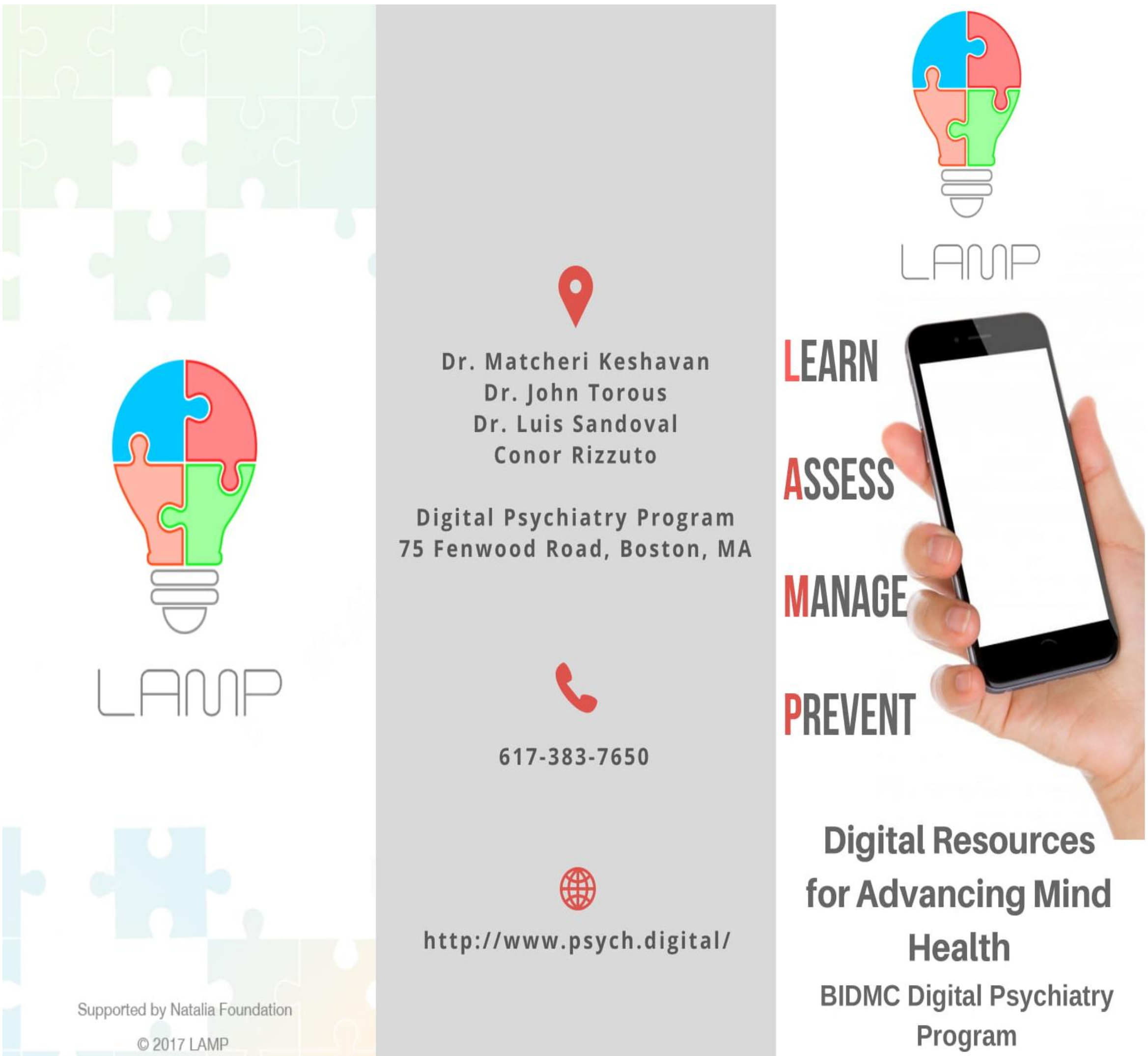
The study had many limitations, including the small sample size, uneven gender distribution and the absence of a control group. In addition to these, the study would be very difficult to reproduce due to the novelty of applications like these. A control group would be useful to have in the future to perform tests such as the receiver operating characteristic to determine the sensitivity and specificity of the application to certain symptoms. The ideal control group for an analysis like this would be college students where the application would act as a measure of school stress levels. The small sample size is being addressed with the next iteration of the smart phone application, called LAMP (Learn, Assess, Manage, Prevent; Appendix 1a and 1b), in which 60 patients will be recruited for the study. Among other benefits, statistical analyses will be more resistant to covariate adjustments and outliers with a larger sample size. LAMP uses *Biewe* for passive data collection.

A study that was recently published in the *Journal of Psychiatric Research* by a team at UC San Diego made a similar effort to validate the use of smartphone applications in the world of healthcare, and specifically to aid in the treatment of early psychosis. Retention and survey completion rates were promising, with 66% of participants remaining in the study for at least 6 months, 77% of the weekly surveys completed and 69% of the daily surveys completed over the course of the study. Participants reported that they would be open to using the application if it was part of their treatment (Niendam et al, 2018). The study involved a similar population of schizophrenic and bipolar patients, as well as a large sample size (n = 76), so it is reassuring that in future studies with similar patients and a large sample size, participation rates have the potential to be even higher than what was seen in this study (57% Android users, 49% Apple users).

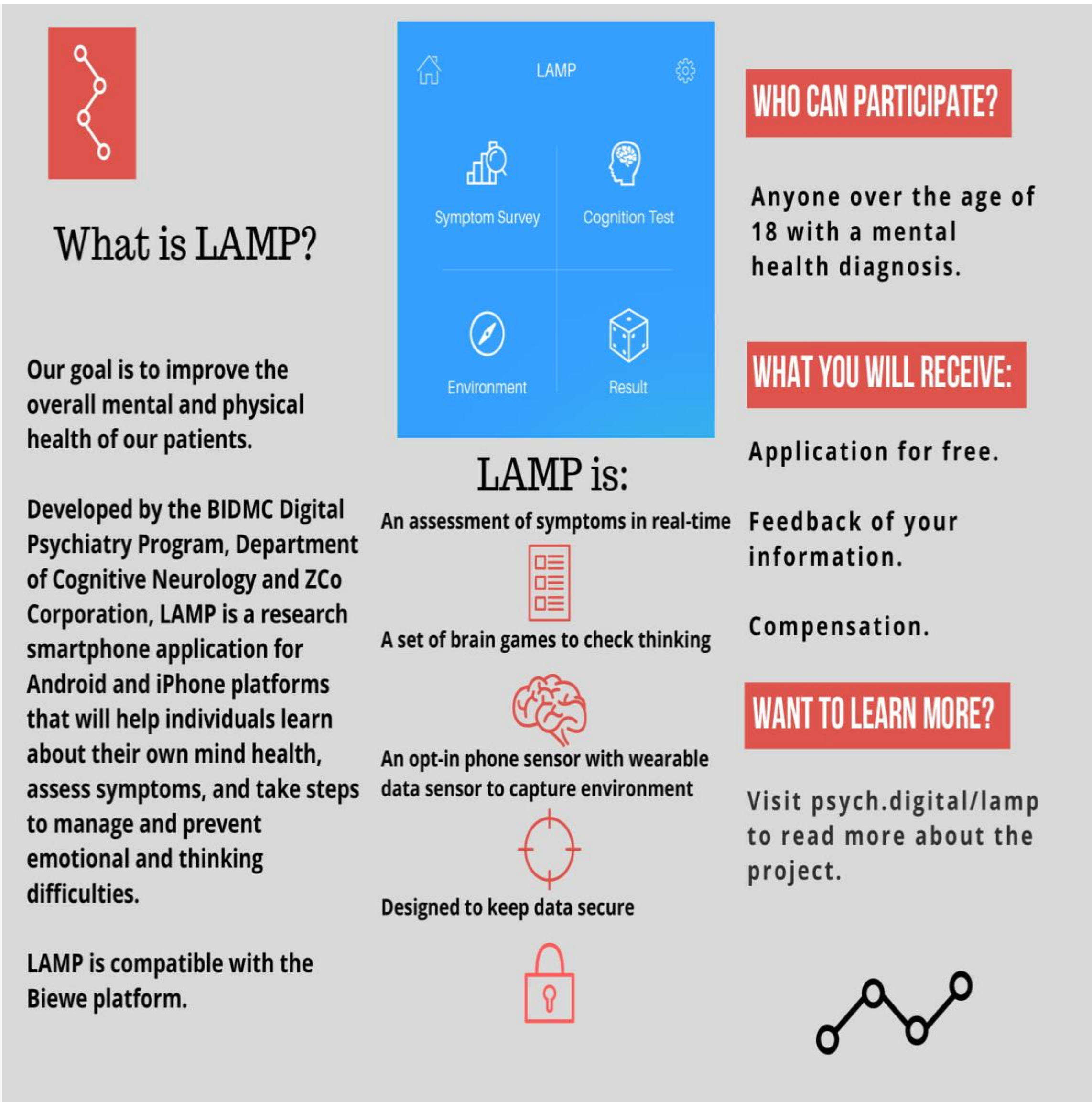
Future studies will want to look into the predictive validity of mobile monitoring applications. One of the objectives of the upcoming LAMP study is to determine reliable predictors of relapse which would be of great benefit to clinical practice. As the study by Niendam et al (2018) pointed out, it will be necessary to arrive at a consensus on how frequently surveys need to be completed for the application to still yield valuable information, since reducing the burden on the patient during long term treatments will be of utmost importance. This study and others like it are not meant to make a case for the removal of clinical scales, but rather to enhance the current state of disease assessment, management and prevention.

APPENDIX

Appendix 1a: Outside of the brochure created for the next iteration of the smartphone application, now called LAMP.



Appendix 1b: Inside of brochure created for the next iteration of the smartphone application, now called LAMP.



LIST OF JOURNAL ABBREVIATIONS

JPR	Journal of Psychiatric Research

REFERENCES

Arsova, Slavica, and Gabriela Kopacheva Barsova. "Patients with Schizophrenia and Social Contacts." *Open Access Macedonian Journal of Medical Sciences*, vol. 4, no. 3, Sept. 2016, pp. 388–91. *PubMed Central*, doi:10.3889/oamjms.2016.084.

Bakdash, Jonathan Z., and Laura R. Marusich. "Repeated Measures Correlation." *Frontiers in Psychology* 8 (2017): 456. *PMC*.

Beiwinkel, Till et al. "Supportive Mental Health Self-Monitoring among Smartphone Users with Psychological Distress: Protocol for a Fully Mobile Randomized Controlled Trial." *Frontiers in Public Health* 5 (2017): 249. *PMC*.

Bersani FS, et al. "Cingulate Cortex In Schizophrenia: Its Relation With Negative Symptoms And Psychotic Onset. A Review Study. - Pubmed - NCBI . Ncbi.nlm.nih.gov. N. p., 2018.

Brambilla, Paolo, et al. "Investigation of Corpus Callosum in Schizophrenia with Diffusion Imaging." *Schizophrenia Research*, vol. 79, no. 2–3, Nov. 2005, pp. 201–10. *PubMed*, doi:10.1016/j.schres.2005.04.012.

Canzian L, Musolesi M. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing. ACM; 2015.

Carbon, Maren, and Christoph U. Correll. "Thinking and Acting beyond the Positive: The Role of the Cognitive and Negative Symptoms in Schizophrenia." *CNS Spectrums*, vol. 19 Suppl 1, Dec. 2014, pp. 38-52; quiz 35-37, 53. *PubMed*, doi:10.1017/S1092852914000601.

Choi, Jung-Seok, et al. "Decreased Caudal Anterior Cingulate Gyrus Volume and Positive Symptoms in Schizophrenia." *Psychiatry Research: Neuroimaging*, vol. 139, no. 3, Aug. 2005, pp. 239–47. *www.psyn-journal.com*, doi:10.1016/j.pscychresns.2004.05.008.

Chue, Pierre, and Justine K Lalonde. "Addressing the Unmet Needs of Patients with Persistent Negative Symptoms of Schizophrenia: Emerging Pharmacological Treatment Options." *Neuropsychiatric Disease and Treatment*10 (2014): 777–789. *PMC*.

Ebdrup, Bjørn H. et al. "Hippocampal and Caudate Volume Reductions in Antipsychotic Naive First-Episode Schizophrenia." *Journal of Psychiatry & Neuroscience :*

JPN 35.2 (2010): 95–104. *PMC*.

Emsley, Robin, et al. "The Nature of Relapse in Schizophrenia." *BMC Psychiatry*, vol. 13, Feb. 2013, p. 50. *PubMed Central*, doi:10.1186/1471-244X-13-50.

Falchook, Aaron D., et al. "Use of Mobile Device Technology to Continuously Collect Patient-Reported Symptoms during Radiation Therapy for Head and Neck Cancer: A Prospective Feasibility Study." *Advances in Radiation Oncology*, vol. 1, no. 2, Feb. 2016, pp. 115–21. *PubMed Central*, doi:10.1016/j.adro.2016.02.001.

Feigenson, Keith A., Alex W. Kusnecov, and Steven M. Silverstein. "Inflammation and the Two-Hit Hypothesis of Schizophrenia." *Neuroscience and biobehavioral reviews* 38 (2014): 72–93. *PMC*.

Firth, Joseph, and John Torous. "Smartphone Apps for Schizophrenia: A Systematic Review." Ed. Gunther Eysenbach. *JMIR mHealth and uHealth* 3.4 (2015): e102. *PMC*.

Fortier, Michelle A. et al. "Pain Buddy: A Novel Use of M-Health in the Management of Children's Cancer Pain." *Computers in biology and medicine*76 (2016): 202–214. *PMC*.

Jee, Haemi. "Review of Researches on Smartphone Applications for Physical Activity Promotion in Healthy Adults." *Journal of Exercise Rehabilitation* 13.1 (2017): 3–11. *PMC*.

Nestor, Paul G. et al. "Dissociable Contributions of MRI Volume Reductions of Superior Temporal and Fusiform Gyri to Symptoms and Neuropsychology in Schizophrenia" *Schizophrenia research* 91.1-3 (2007): 103–106. *PMC*.

Niendam, Tara A., et al. "Enhancing Early Psychosis Treatment Using Smartphone Technology: A Longitudinal Feasibility and Validity Study." *Journal of Psychiatric Research*, vol. 96, Jan. 2018, pp. 239–46. *ScienceDirect*, doi:10.1016/j.jpsychires.2017.10.017.

Ong, Stephanie W., et al. "Integrating a Smartphone–Based Self–Management System into Usual Care of Advanced CKD." *Clinical Journal of the American Society of Nephrology : CJASN*, vol. 11, no. 6, June 2016, pp. 1054–62. *PubMed Central*, doi:10.2215/CJN.10681015.

Palmier-Claus, Jasper E et al. "The Feasibility and Validity of Ambulatory Self-Report of Psychotic Symptoms Using a Smartphone Software Application." *BMC Psychiatry* 12 (2012): 172. *PMC*.

Picchioni, Marco M, and Robin M Murray. “Schizophrenia.” *BMJ :British Medical Journal* 335.7610 (2007): 91–95. *PMC*.

Scheepers, F. E., et al. “Effect of Clozapine on Caudate Nucleus Volume in Relation to Symptoms of Schizophrenia.” *The American Journal of Psychiatry*, vol. 158, no. 4, Apr. 2001, pp. 644–46. *PubMed*, doi:10.1176/appi.ajp.158.4.644.

Takayanagi, Mizuho et al. “Reduced Anterior Cingulate Gray Matter Volume and Thickness in Subjects with Deficit Schizophrenia.” *Schizophrenia research* 150.0 (2013): 484–490. *PMC*.

Uematsu, Masayuki, and Hisanobu Kaiya. “The Morphology of the Corpus Callosum in Schizophrenia: An MRI Study.” *Schizophrenia Research*, vol. 1, no. 6, Nov. 1988, pp. 391–98. *ScienceDirect*, doi:10.1016/0920-9964(88)90020-5.

Wang, Lei et al. “Abnormalities of Cingulate Gyrus Neuroanatomy in Schizophrenia.” *Schizophrenia research* 93.1-3 (2007): 66–78. *PMC*.

Xiao, Yuan, et al. “Similar and Different Gray Matter Deficits in Schizophrenia Patients and Their Unaffected Biological Relatives.” *Frontiers in Psychiatry*, vol. 4, 2013. *Frontiers*, doi:10.3389/fpsyt.2013.00150.

Yoshimura, Shinpei, et al. “Rostral Anterior Cingulate Cortex Activity Mediates the Relationship between the Depressive Symptoms and the Medial Prefrontal Cortex Activity.” *Journal of Affective Disorders*, vol. 122, no. 1-2, 2010, pp. 76–85., doi:10.1016/j.jad.2009.06.017.

CURRICULUM VITAE

[redacted]

[redacted]

[redacted]