High frequency assessment of mood and cognition in major depressive disorder using the Apple Watch

F Cormack1,2, M McCue3, J Schuster4, N Tapiotakis1,2, E Glazer1,4, E Panagopoulos1,4, T van Schaik1,4, B Fehnert1,4, J King1,4, JH Barnett1,2,5

1 Cognition Kit, Cambridge UK; 2 Cambridge Cognition, Cambridge UK; 3 Takeda Pharmaceuticals USA, Deerfield IL; 4 Citi Group, London UK; 5 University of Cambridge Department of Psychiatry, Cambridge UK

This work was sponsored and funded by Takeda Pharmaceuticals.

Background

Patients with major depressive disorder (MDD) frequently experience cognitive problems such as inattention and memory complaints (Porter et al., 2007; Hammal & Antal, 2010). These are often under-recognized and may affect patients’ ability to accurately self-report symptoms (Baume et al., 2010; Gupta et al., 2013). High frequency, near-patient testing may aid communication between patients and clinicians by overcoming challenges such as recall bias. This study aimed to assess the feasibility and accuracy of daily assessments of cognition and mood implemented on an Apple Watch in participants with MDD.

Methods

Participants

Thirty adults aged 19-63 with mild to moderate single or recurrent Major Depressive Episode prescribed antidepressant monotherapy were enrolled (baseline characteristics in table 1).

Daily assessments

Brief cognitive (n-back) and mood assessments were administered through the Cognition Kit application on Apple Watch in a single-arm, 6-week observational study (Figure 1).

Compliance with daily assessment

Compliance results are shown in Figure 2 and 3, plotting individual participant (x-axis) by daily compliance during the 6-week study period (n=axis), yellow squares show daily completed assessments, black where they were missed.

Daily mood assessments:

Self-reported mood was available to complete once per day. 50% of participants achieved 100% compliance, completing all assessments; overall compliance was 94.6%.

Trouble concentrating on things (e.g. newspaper, TV)

Lack of interest or pleasure in doing things

Feeling down, depressed or hopeless

Top watch image: example of cognitive assessment; bottom: mood assessment

Additional full-length assessments

Up to 5 additional assessments were completed.

Compliance with daily assessment

Results

Patient Characteristics

Table 1: Participant demographic characteristics at baseline

<table>
<thead>
<tr>
<th>Gender (Female)</th>
<th>N</th>
<th>%</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>16</td>
<td>55</td>
<td>3.05</td>
<td>1.02</td>
<td>1.44</td>
<td>6.8</td>
</tr>
<tr>
<td>Female</td>
<td>28</td>
<td>45</td>
<td>3.01</td>
<td>1.01</td>
<td>1.56</td>
<td>6.7</td>
</tr>
</tbody>
</table>

Correlations with full-length assessments

Table 3: Significant Pearson correlation coefficients between aggregate daily mood assessments and all full length PRO measures

<table>
<thead>
<tr>
<th>Total mood score</th>
<th>Feeling down, depressed</th>
<th>Trouble concentrating</th>
<th>Lack of interest or pleasure</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRO-9</td>
<td>0.80</td>
<td>0.76</td>
<td>0.64</td>
</tr>
<tr>
<td>HAMA</td>
<td>0.68</td>
<td>0.67</td>
<td>0.65</td>
</tr>
<tr>
<td>LS-19</td>
<td>0.40*</td>
<td>0.47</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 3: Significant Pearson correlation coefficients between aggregate daily mood assessments and all full length PRO measures

Performance on n-back task (inset graph) shown on the y-axis, study day on a x-axis, learning slope and intercept shown in colour for each participant. Performance was modelled using longitudinal mixed effects model with daily average n-back score (d-prime) as response variable, a fixed effect of study day, random effect of participant with random intercept and random slope.

Correlations with daily cognitive function

n-back performance improved over time, with considerable between-subject variance in the intercept and slope of the learning curves. Few ceiling effects were observed. Figure 5 shows individual participant scores and learning trajectories from the mixed-effects model. Daily n-back performance correlated significantly with CANTAB working memory (SWM) and attention (RVP) tests, and PROs:

- Working memory test (SWM): between errors (measuring working memory capacity) correlated daily n-back performance (mean d-prime = -46, p<0.01; intercept: -51, p=0.01). Search strategy correlated daily n-back performance (mean d-prime: r=-0.37, p<0.05; intercept: r=-0.40, p<0.05).

- Attention test (RVP): Sensitivity to a target correlated daily n-back performance (mean d-prime: r=0.50, p<0.01; intercept: r=0.47, p<0.01). Mean response latency correlated with daily n-back performance (mean d-prime: r=0.40, p<0.05; intercept: r=0.42, p<0.02).

- PRO-9 score correlated with mean d-prime: r=0.38, p<0.05.

Qualitative Insights

Participants completed qualitative entry and exit interviews.

Some found participants achieving compliance easier than others, depending on how the study satiates fit in with their daily routines, moods and lifestyles.

Intrinsic and extrinsic motivators worked together to encourage engagement at different stages of the study. Participants reported that their own mood directly impacted on motivation and compliance, and that it was easier to complete tasks in private environments. Many reported that completing daily cognition and mood tasks became a daily challenge that contributed to a sense of achievement.

Participants reported that specific features of the app and device contributed to their engagement, many found the Apple Watch itself to be a motivating factor.

Conclusions

- Near-patient testing using wearable devices is feasible and well-tolerated by patients with depression, with good-to-excellent task completion achieved.

- Correlations with patient reported outcomes and objective measure was shown, providing a novel, patient-centric methodology for frequent assessment of a range of symptoms.

- Rapidification in larger sample is required.