Background

The early stages of many neurodegenerative conditions, including MCI and prodromal AD, are characterised by a fluctuating course, which impacts functioning, but is difficult to characterise on the basis of infrequent laboratory or clinical assessments. Detecting and characterising these fluctuations may provide a more accurate and stable baseline for clinical trials of disease-modifying treatments, and ultimately enable timely, tailored interventions.

Wearable devices and mobile phones are equipped with increasingly sophisticated sensors and processing capacity. This technology generates large volumes of multidimensional data, which can increasingly be linked to changes in symptoms and functional status.

In this study we describe the development of cognitive testing and mood data collection capabilities on a consumer wearable device, which also allows the measurement of physiological parameters.

Methods

Participants

10 (4 females) participants, aged 24–55 were recruited to this two-week feasibility study. All were educated to at least university level, and all were in full or part time employment.

Wearable Device

- The device used was the Microsoft Band 2. We extracted data on Heart Rate, RR-interval providing heart rate variability data, and skin conductance response. Together these provide an indication of autonomic arousal.
- Data from these sensors was streamed via Bluetooth to the participant’s mobile device, which in turn connected to a cloud-based study management system, which scheduled sensor data collection, and behavioural testing (Figure 1).

Cognitive Testing

- Wearable-based cognitive testing was carried out using a 2-back memory paradigm (Figure 2). This task taps several aspects of cognition, including attention, memory updating and working memory. Each test took 1 minute to complete.
- Reaction time and accuracy were recorded, and a measure of accuracy taking into account both hits and false alarms was computed (d’). At the end of each test, participants received feedback regarding their accuracy.

Mood Rating

- Emotional state was probed immediately after cognitive testing (Figure 3).
- Participants were asked to select the icon which conveyed their current emotion, and then rate the intensity of this emotion. Verbal labels were provided alongside each pictogram.

Study Design

Participants were provided with an MS Band 2 which was paired with their iPhone. Sensor data was recorded continuously between 9:00 and 19:00, with cognitive and mood testing once an hour over the study period. When a test was scheduled the participant received an alert. If a participant was not able to complete the test they ignored the alert. The alert was repeated twice more at 5 minute intervals.

Validation tasks were administered daily in the evening. These include four measures from the Paired Associates Battery (Figure 4) to assess the correspondence between n-back and other measures of cognition, and the Positive and Negative Affect Schedule (PANAS) as a measure of daily mood. Participants provided feedback on the user experience which was analysed to extract key themes.

Results: Cognitive performance

Individual differences in the distribution of n-back performance were observed (A), with three participants consistently reaching ceiling levels on the 2-back task.

- A mixed quadratic growth model was applied to the data, modeling learning effects over time. This allowed for random intercepts, and slopes, with fixed quadratic terms. Individual performance curves are plotted in Figure 4C.

Predictors of the intercept and slope of these curves are plotted in relation to age (C).

- Spearman correlations with CANTAB scores are shown in D. The magnitude of the correlations varied from 0.04 (FVP and learning slope) to 0.60 (FVP false alarm rate and learning slope).

The presence of learning effects, and significant interactions between performance, age and measures of cognition support the viability of brief, repeated n-back performance as an indicator of cognitive function.

Results: Compliance with data collection

There was good compliance with data collection, with more than 60 n-back trials being collected over the course of each day.

- More testing sessions were completed on week days than weekends (E). Hours coinciding with commuting (9 am and 6 pm) produced fewer tests, compared with the rest of the testing times (F).

The number of push notifications sent to the band before a participant responded is shown in G. The majority of responses were received after one notification. More were received in some instances, e.g. when the band was outside of Bluetooth range.

Sensor coverage was assessed in the hour preceding cognitive assessment to evaluate the density of measurement and the length of continuous recordings achieved. Sensor coverage was generally acceptable, and best in the periods immediately preceding cognitive testing (H).

Compliance in context: qualitative insights

Qualitative analysis of participant’s experience revealed themes accounting for adoption, compliance and perceived impacts on data quality:

- Accounting for context - cognitive performance is impacted by distractions in real-world environments and the varying strategies people choose to manage them.
- Adapting to routines - changes in routine (i.e. weekdays to weekends) shift individual compliance patterns over time and impact practical challenges to test completion (e.g. phone and band charging).
- Giving meaning - giving feedback on performance, allowing participants to improve with practice, and demonstrating an aim to provide societal value contributes to adoption, motivation and compliance.