Author’s response to reviews

Title: The feasibility and validity of ambulatory self-report of psychotic symptoms using mobile phones

Authors:

Jasper E Palmier-Claus (Jasper.Palmier-Claus@manchester.ac.uk)
John Ainsworth (John.Ainsworth@manchester.ac.uk)
Matthew Machin (Matthew.Machin@manchester.ac.uk)
Christine Barrowclough (Christine.Barrowclough@manchester.ac.uk)
Graham Dunn (Graham.Dunn@manchester.ac.uk)
Emma Barkus (ebarkus@uow.edu.au)
Anne Rogers (Anne.Rogers@manchester.ac.uk)
Til Wykes (til.wykes@kcl.ac.uk)
Shitij Kapur (Shitij.Kapur@kcl.ac.uk)
Iain Buchan (lain.Buchan@manchester.ac.uk)
Emma Salter (Emma.Salter@student.manchester.ac.uk)
Shôn Lewis (Shon.Lewis@manchester.ac.uk)

Version: 3 Date: 19 July 2012

Author’s response to reviews: see over
We would like to thank the reviewers for their helpful and constructive comments. These were extremely useful and allowed us to identify areas in the paper where adjustments, expansion or further clarification were beneficial. We feel the paper has been improved overall as a consequence. All reviewer comments are presented below in black. Below each comment is our response to these comments. Quotations from the text are italicised and page numbers are provided where relevant.

Reviewer's report
Title: A new mobile assessment technology for psychosis.
Version: 2 Date: 21 June 2012
Reviewer: David Kimhy
Reviewer's report:
Overall Review:
The authors compared use of mobile devices and traditional assessments of psychotic and other symptoms among individuals with schizophrenia and related disorders. The manuscript focuses on an important and timely topic, namely the emerging use of mobile assessment technology to conduct "real-world" ambulatory assessment of psychotic symptoms in individuals. While the focus of the manuscript on use of wireless technology is exciting, my enthusiasm was somewhat diminished due to the fact that the actual wireless component of the technology was not utilized in this study. Furthermore, even if this component was included in the study, the authors’ failure to provide context from studies that used earlier technologies (i.e., PDAs), makes it is difficult to evaluate the potential advantages/disadvantages of their technology. There are a few additional issues that if addressed, would improve the manuscript:
Major Revisions:
1. Page 4, line 4 - The authors stated that compared to previous studies that utilized PDAs, their use of mobile phones to collect ambulatory data is advantageous, as “people are accustomed to carrying and recharging [such devices] … and are often familiar with the technology and interface.” The authors also claim that PDAs “are limited by their short battery life and difficult user-interface.” Can the authors present data indicating that mobile phones have superior battery life compared to PDAs? Likewise, if the ClinTouch software is designed to work on numerous Android-supported phones, and each phone would potentially have a unique battery-life profile, how can such a broad claim be made? Regarding the interface and comfort of use claim, a quick inspection of the ClinTouch interface indicates that it is quite similar to those used in previous studies using PDAs. What are the advantages in comfort of use? The authors need to provide clear data to support these claims.

The poor battery life and difficult interface is a limitation of PDAs raised by Hektner and colleagues (2007). However, the reviewer makes a fair point that this is a bold claim given the variety of different handsets available. We have therefore removed this sentence from the introduction and added a reference to support the claim that individuals are becoming increasingly familiar with SmartPhone technology.
An appropriately enabled mobile phone may have the advantage that people are accustomed to carrying and recharging it and are often familiar with the technology. Software applications are also easily uploaded to participants’ own SmartPhones ensuring that the individual does not have to carry with them an additional device. In a recent Ofcom report in the United Kingdom, 27% of adults and 47% of teenagers currently owned a SmartPhone [16]. With advances in mobile phone technology PDAs are becoming increasingly obsolescent.

2. The authors correctly state that PDAs “are offline systems and whilst the data is collected in the real world it cannot be assessed until brought into the laboratory/clinic and downloaded.” The authors suggest that their methodology is therefore advantageous. However, they need to state clearly what are the practical, everyday advantages of having “live” data from participants. Specifically, as data for many analyses (i.e., compliance rates & mean symptom ratings) cannot be computed until the ambulatory assessment is completed, what are the actual advantages of obtaining the data a few hours/days earlier?

The advantages of having real-time uploading of information lie in its clinical application. For example, it may help to alert clinicians to symptom worsening or relapse, which could help to facilitate earlier intervention. The information would not necessarily need aggregating/averaging in order to do this; clinicians could access automatically updated plots or graphs showing their patient’s symptoms progression over time (at each individual time-point). Additionally, the system could trigger alerts to the clinician should an individual’s symptoms start to deteriorate. We have attempted to make the clinical applications of the technology clearer in the text.

PDAs are offline systems and whilst the data is collected in the real world it cannot be assessed until brought into the laboratory/clinic and downloaded. Assessing data in vivo is desirable in that it could help to facilitate earlier and more immediate intervention, which in turn could help to reduce relapse, self-injury and the need for unscheduled acute care. Automated and personalised feedback could help clinicians to devise and review treatment strategies prior to consultation allowing for more effective care.

3. The authors claim that their use of mobile phones is advantageous compared to PDAs. However, given that the wireless connectivity was not used as part of this study (page 10, line 14), it appears that the present investigation is not much different than previous studies that utilized PDAs, perhaps with a slightly “smarter” Android interface. The utilization of any new technology is often accompanied by technical problems that need to be overcome. Until the authors complete a study in which the wireless component of the their devices runs satisfactorily and problem-free, the authors’ claims of validating a “new” technology appear premature.

We feel that this technology is novel given that it runs on a different type of device, which has the capacity to have wireless connectivity. Additionally, the software itself differs from other technologies (e.g. the delusion menu, administrator’s page, alarm volume etc).

4. In the Manuscripts’ title, the authors characterize the technology they are using
as “new”. However, this characterization is somewhat misleading as PDAs and similar technologies have been used to study individuals with schizophrenia for at least half a dozen years (See Kimhy et al., 2006, 2010; Granholm et al., 2008, 2112; Ben-Zeev et al., 2011; Swendsen et al., 2011 to mention a few; See Kimhy et al., 2012 for review). Likewise, preliminary reports of using such technologies as part of treatment of individuals at high risk for psychosis have been published as well (Kimhy & Corcoran, 2008). Overall, the manuscript may benefit from incorporating information from previous related studies, which will allow comparison of the advantages, and disadvantages of the presented technology compared to previous ones.

We have now made reference to the levels of compliance seen in past PDA studies in order to give the reader a better understanding of the general literature. In addition, we have adjusted the title to remove the word “new”, and amended the abstract:

‘Over the past decade Personal Digital Assistants (PDAs) have been adapted for self-report symptom monitoring in individuals with severe mental illness [12]. Studies evaluating PDAs have shown low rates of drop-out in community dwelling individuals with psychotic disorders [13-15]. For example, Granholm and colleagues [13] found that 87% of patients were compliant to PDA based momentary assessment as defined as completing at least four out of 28 data-points. Other studies have observed similarly low rates of drop-out when using more conservative definitions of compliance (eight out of 28 data-points) [15].’ p6

5. Can the authors comment about the relatively lack of paranoid delusions among participants (as listed on page 9) and whether this reflects selective recruitment (i.e., fewer individuals with paranoia agreed to participate and use their personal mobile phones.)

We created an item in the delusions menu to assess the specific belief that there was a conspiracy against the individual. We also created a suspiciousness momentary assessment scale, which assessed feelings of paranoia more generally. There were a number of patients who reported paranoia on the latter scale, but did not have the specific belief that there was a conspiracy against them. Indeed we observed a wide range of paranoia scores both on the PANSS (1-6) and the paranoia momentary assessment scale (see Table 2) suggesting that the sample did indeed experience paranoid delusions.

6. The authors state that they do not know how many participants were approached (and refused) to participate in the study. This is a major limitation of the study and should be highlighted in the limitation section of the manuscript. While the authors stated that 82% of participants were compliant, given the unknown (50%?) acceptability of the study and methodology to participants, broad claims that the methodology is feasible and acceptable appear premature (Page 16, line 3). This is also true regarding claims that ClintTouch may one day even replace traditional assessment (Page 20, Line 3). As stated earlier, a better characterization of the sample regarding socioeconomic status, education and/or reading ability would potentially help such claims.
We agree this is a limitation which needs to be addressed. We have provided an estimation of the refusal rate and have mentioned it as a limitation in the discussion. We have also have now provided information about those participants who were referred to the research team, but were subsequently ineligible or declined to take part:

‘Initial verbal approach to participate was made by a member of the clinical care team and about 50% of those approached declined to take part. Of the 51 patients who agreed to be contacted about the study and had their contact details passed on to the research team, four subsequently declined, two were ineligible and one could no longer be contacted.’ p15

We have pointed to a lack data on socioeconomic status and reading ability as a disadvantage of our analyses. The line that this technology might replace traditional assessments has been removed.

‘It should be noted that although compliance was high in this study rates of refusal to initially take part could not be assessed. Furthermore, socioeconomic status and reading ability were not considered, which may have predicted levels of non-compliance.’ p20

Years of education has been added to Table 1. This seems roughly equivalent to other studies in this area (e.g. Ben-Zeev and colleague’s (2011) sample had a mean years of education of 12.4).

<table>
<thead>
<tr>
<th></th>
<th>Acute (n=12)</th>
<th>Remitted (n=12)</th>
<th>UHR (n=12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years of education, mean (SD)</td>
<td>11.7 (2.5)</td>
<td>12.0 (3.0)</td>
<td>12.6 (1.6)</td>
</tr>
</tbody>
</table>

7. The authors describe evaluating compliance as one of the primary aims of the study. However, as stated earlier, they did not review information about compliance in previous studies, which would have allowed comparisons (Kimhy et al., 2012 provided a summary of such findings).

We have now compared our results to the compliance rates seen in past momentary assessment studies. The findings are relatively similar.

‘In compliant individuals, the number of assessment occasions was relatively high and similar to past momentary assessment research using PDAs in this population. For example, Swendsen and colleagues [13] observed an identical completion rate of 72% of all data-point completed, whereas Granholm and colleagues [11] found this to be 69%.’ P19

Minor Revisions:
1. Page 2, Line 1 – “Schizophrenia is one of the most prevalent forms of mental illness.” Given that the general prevalence of schizophrenia is about 1%, characterizing schizophrenia as one of the most prevalent mental disorders is an overstatement.
We have changed this to:

‘Schizophrenia is often hugely distressing and disabling to the individual [1], with an associated cost in the United Kingdom of around 6.7 billion pounds each year [2].’ P5

2. Page 4, Line 13 – the “and” appears in error in “The objective of this study was to and evaluate.”

We have removed the error.

3. There are discrepancies between the aims as described in the abstract, the aims listed on page 4, and the hypotheses listed on page 5. Please address.

We have made the abstract, objectives section and hypotheses section of the introduction more consistent. Due to our word count we have limited the abstract to our primary hypotheses (see p4)

4. The number of participants listed on page 5 is different from the numbers listed in the Abstract, the Results section and Table 1. Please correct.

We have corrected this.

5. It is unclear what was the minimal completion rate necessary for subject’s data to be included in the analyses (33%, top of page 12 vs. 67% bottom of page 12). Please clarify. Also, how was this standard determined?

Individuals had to complete at least 33% of entries in order to be deemed complaint with the methodology consistent with the criteria set by the Maastricht group (Myin-Germeys et al., 2001, 2005 etc). We have made note of this in the methods section.

‘Eligible participants were prospectively recruited into the three groups until each group contained 12 subjects who had managed to complete at least 33% of the 42 data entry points possible during the six consecutive days of testing in accordance with momentary assessment past studies [19]’ p8

6. Can the authors provide information about the socioeconomic status, education and/or reading ability of the participants? As these variables may have an impact on the results, including them may help determine the generalizability of the findings. If such information is not available, this issue should be addressed in the limitations of the manuscript.

We have added this as a limitation of the study.

‘Furthermore, socioeconomic status and reading ability were not considered, which may have predicted levels of non-compliance.’ p20
7. Many individuals with schizophrenia live on limited incomes. Given that the ClinTouch software is installed on the participants’ personal phone – are participants responsible for paying for the additional communication usage associated with the software?

In the future we will be setting up a number that is free to text information to; therefore the transfer of information will not incur any costs to the patient.

Discretionary Revisions:
1. Can the authors provide data regarding the participants’ response rate over the course of the seven days? Such data may provide information regarding potential declines in the response rate (day 1 vs. day 6) and hint regarding the feasibility of using this system for monitoring symptoms over longer periods of time.

We are hoping to cover this in a later publication and therefore would rather not include the results of this analysis in the current paper.

Level of interest: An article whose findings are important to those with closely related research interests
Quality of written English: Needs some language corrections before being published
Statistical review: Yes, but I do not feel adequately qualified to assess the statistics.
Declaration of competing interests: I declare that I have no competing interests

Reviewer’s report
Title: A new mobile assessment technology for psychosis.
Version: 2 Date: 15 June 2012
Reviewer: Bryan McCormick
Reviewer’s report:
- Major Compulsory Revisions
  1. The authors noted that two (sub)sets of questions from the PANSS and CDS were created and alternately presented to participants. A presentation of how the sets were created and if they were intended to be parallel or complementary is necessary.
  2. Based on the description in the statistics section, it appears that individual items were summed across the two sets of response items to create scores. If I understand this correctly there would appear to be some methodological concerns about summing items that are not referencing the same time period. For example, as there are 4 anxiety items what bias is introduced if items 1 & 2 reference time period 1 in day 1 and items 3 & 4 reference time period 2 in day 1? As with clarification of instrumentation in #1 above, clarification of statistical procedures is necessary as well.
In response to 1 & 2:

We did not split items within a symptom scale across sets (e.g. all anxiety items were always assessed at the same time-point). We have attempted to clarify this and provide a rationale for how the sets were created in the methods section:

‘In order to reduce the length of time taken to complete the items, these were divided into two sets, which alternated across time-points. Guilt, hopelessness, depression, social withdrawal, conceptual disorganisation, excitement and hallucinations were assessed in set one, whereas anxiety, grandiosity, hostility, somatic concern, guilty ideas of reference, paranoia and delusions were assessed in set two. The allocation of scales to the two sets was based on the need to assess overlapping symptom domains (e.g. paranoia and delusions) at the same time-point and to keep the number of items balanced. Some of the self-report scales were branched so that the use of certain items was contingent on the participant’s previous response. The stem and branching questions related to the constructs measured on the PANSS and the CDS items, while being compatible with self-report. Thus, 15 to 30 questions were presented in set one and 11 to 31 questions were presented in set two.’ P11

3. In addition, one of the hypotheses related to the internal consistency (alpha) of the self-report items; however, no findings from a reliability analysis were reported or discussed.

The alphas were placed in table 2, but we had not made reference to them in the text. We have rectified this in the methods and discussion sections.

‘The internal consistency and instability of the scales
As can be seen in Table 2, the alpha scores for each of momentary assessment scales were high suggesting good internal consistency.’ P17

‘All of the self-report scales also showed good internal consistency.’p21

- Minor Essential Revisions
1. Source(s) for accessing participants needs to be clarified. Were they drawn from a health service, etc?

We have added information on the services participants were recruited from in Table 1:

<table>
<thead>
<tr>
<th>Service recruited through, n</th>
<th>Acute (n=12)</th>
<th>Remitted (n=12)</th>
<th>UHR (n=12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMHT*</td>
<td>0</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>Early intervention</td>
<td>2</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>Inpatient ward</td>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Early detection</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>
2. There is some confusion in the reporting of findings as to whether a 67% compliance rate was used or if a 33% compliance rate was used. The authors cite a 67% rate of compliance, but then list a minimum of 14 entries among the acute group. Given a stated 6 signals/day over 7 days, 14 entries corresponds to a 33% rate of compliance.

Yes, participant had to complete at least 33% (N=14) of all possible entries over the week of sampling in order to be deemed compliant with the methodology. We have corrected this error in the text.

3. I believe that the authors need to provide a greater acknowledgement of the challenges of this data collection procedure in accessing psychiatric symptoms. For example, they do not adequately address the drop out rate in the acute group. They also do not discuss the issue of adequacy of responses. Are 14 data points (the minimum identified) enough for accurate identification of psychiatric symptoms?

We have attempted to highlight the challenges of this data collection procedure in the discussion section.

‘The number of individuals dropping out of the study was relatively low across remitted and UHR samples, although slightly elevated in acute patients, where a third of individuals were non-compliant. This may explain the finding that positive symptoms significantly predicted non-compliance to the procedure. This supports the notion that momentary assessment is a relatively demanding approach, to which more symptomatic and chaotic patients may have difficulty in remaining compliant [21]. Thus, in acute settings it may be beneficial to adapt the momentary assessment procedure (e.g. sampling rate, item number) to individual’s preferences and needs, or use an alternative method of assessment.’ P19

- Discretionary Revisions
1. I don’t know that the title accurately reflects the substance of the work. While this is clearly related to mobile assessment technology, it is more directly related to psychiatric symptoms, as opposed to the generic "psychosis."

This technology was specifically designed for psychosis even though it does assess a range of related symptoms. Therefore, we request that we leave the title as it stands.

2. Indication of the distribution across groups of the 12 cases used to establish inter-rater reliability on the PANSS would be helpful to the reader.

These were 4 UHR, 5 acute and 3 remitted. We have now stated this in the text:
Excellent inter-rater reliability was demonstrated with an independent rater rating 12 (5 acute, 3 remitted and 4 UHR) of the audio-recorded PANSS interviews (Spearman’s correlations, PANSS positive subscale, rho = .91; negative, rho = .82; global, rho = .81; total, rho = .79)." p9

3. Although I believe that the authors have adequately discussed the issue of sensitivity, they might also consider that their technique may have contributed to findings. At least some research has shown that people with schizophrenia fail to maintain an emotional response (Kring et al. 2011). As the technique required the aggregation of experience over 2 hours (on average) a loss of emotional response may have occurred.

This is a very interesting reference although we are not sure whether the loss of a prolonged emotional response would necessarily lead to poorer recall.

4. Table 3 represents figures that are a mean of means (across all participants). As such, they would have their own indicators of variability that should be reported.

We have now included the between subject SD for both instability metrics in Table 3.

Level of interest: An article whose findings are important to those with closely related research interests
Quality of written English: Acceptable
Statistical review: No, the manuscript does not need to be seen by a statistician.
Declaration of competing interests: I declare that I have no competing interests