PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (http://bmjopen.bmj.com/site/about/resources/checklist.pdf) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

<table>
<thead>
<tr>
<th>TITLE (PROVISIONAL)</th>
<th>Evaluating the Attitudes of Mental Health Professionals Towards Trials of MDMA: A Randomized Vignette Trial</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUTHORS</td>
<td>Wright, Dean; Colagiuri, Ben; Glozier, Nick</td>
</tr>
</tbody>
</table>

VERSION 1 – REVIEW

| REVIEWER           | Bright, Stephen  
|                   | Edith Cowan University                                 |
| REVIEW RETURNED    | 31-May-2022                                               |

| GENERAL COMMENTS   | This paper provides empirical data on mental healthcare professional's attitudes towards MDMA-assisted psychotherapy. The introduction is short and sharp, and the methods adequately describe the study procedures, although I am not sure that all of the supplementary materials included are required for publication. I am curious as to the benefits of this design over a within-subjects design using a vignette for the control condition that was more closely aligned with currently available pharmacotherapies such as SSRIs. Such a within groups design would have controlled for the significant differences in demographics between the two groups and would have allowed for stronger conclusions to be drawn, particularly given the correlations found between age and objection to MDMA. I am also confused as to the p value of 0.067 for this analysis compared with 0.031 for the control given the higher r value for MDMA. And yet the objection to the control is not marked in Table 2 as being significant despite it being reported as <0.05 in the manuscript text. In addition to the articulating the rationale for the between groups design and the reporting issues noted above, some minor issues that also need addressing: 

Page 3, line 8 AP – cite in full this first time it is mentioned 

Page 5, line 29 – MDMA research has not used an active placebo (though it should!). Please rephrase to increase clarity. 

Page 5, line 47 – MDMA is also going to have inflated effect sizes due to the issue of inadequate blinding 

Page 9, line 13 - I’m not sure what the first sentence of the planned analysis means: “Experimenters were blinded prior to conducting statistical analyses”, perhaps reword for clarity? |

| REVIEWER           | Abdin, Edimansyah  
|                   | Institute of Mental Health, Research Division                     |
| REVIEW RETURNED    | 02-Aug-2022                                               |

| GENERAL COMMENTS   | Thanks for giving opportunity to review this manuscript. This is first |
study to experimentally assess bias in MH professional attitudes towards the emerging field of psychedelic psychotherapy, particularly MDMA-AP. The study found there is no bias against MDMA-AP, compared to a trial using an unknown, neutrally labelled pharmacotherapy. I have few comments especially in the method section for authors consideration. I was not clear why randomization based on professions were chosen and why effect size based on minimal clinically difference were chosen when calculating the sample size. Since the sample size was split into two group and further stratified by professions, I was wondering how authors ensure that the sample size is equal across these 4 strata when data collection was conducted? It is also not clear if the sample size calculation has taken into account whether the power is sufficient for these subgroup. It would be good if authors can provide any reference and justify these approach. In statistical analysis section, authors should clarify whether normality test had been done before using parametric tests such as ANOVA etc were used. Although randomization was used, the group seems to be different based on age and experience. It would be good if authors can comment on this finding and it might affect their randomization approach and significant findings.

**VERSION 1 – AUTHOR RESPONSE**

**Reviewer 1 – Dr. Stephen Bright, Edith Cowan University**

Comments to the Author:

This paper provides empirical data on mental healthcare professional’s attitudes towards MDMA-assisted psychotherapy. The introduction is short and sharp, and the methods adequately describe the study procedures, although I am not sure that all the supplementary materials included are required for publication.

2) I am curious as to the benefits of this design over a within-subjects design using a vignette for the control condition that was more closely aligned with currently available pharmacotherapies such as SSRIs.

Such a within groups design would have controlled for the significant differences in demographics between the two groups and would have allowed for stronger conclusions to be drawn, particularly given the correlations found between age and objection to MDMA.

**RESPONSE:**

The reviewer raises two valid questions here and there are strengths and weaknesses to both approaches.

1) The use of a neutrally labelled, unknown research compound as a control enabled us to assess for differences in attitudes specific to the name “MDMA-assisted psychotherapy” and attempt to minimise the potentially large confound of an attitude to novel as opposed to familiar medications. This was chosen at the sacrifice of potential ecological validity of a commonly used compound, such as an SSRI or SNRI. We have included this point in our discussion.

2) Regarding the choice of a between-groups comparison over a within-groups design, Dr Bright is correct – this would likely have reduced some of the variability in our comparisons, but at the expense of removing participant blinding. When considering our design, we had two primary concerns with a within-groups design: 1) By having participants read both vignettes, they effectively become unblinded to the study design, potentially biasing their answers. We wanted to ensure that those in the control group were unaware the study was about MDMA.
2) The longer the intervention becomes (with two vignettes & questionnaires), the less likely participants are to complete the study. By doubling the intervention time, we likely would have lost participants.

3) I am also confused as to the p value of 0.067 for this analysis compared with 0.031 for the control given the higher r value for MDMA. And yet the objection to the control is not marked in Table 2 as being significant despite it being reported as <0.05 in the manuscript text.

RESPONSE:

There appears to be a misunderstanding regarding these results. There are two different types of analysis being reported addressing two different questions.

The first questions assess the correlations between a) objection to participate in any pharmaceutical trial and objection to participate in the control trial (r = .285; p = .067); and b) objection to participate in any pharmaceutical trial and objection to participate in the MDMA trial (r = -.106, p > .05).

The second question compares whether these two correlations above are significantly different (which they are: Fisher r to z: d = .40, p = .031) using a Fisher r to z comparison of the two correlations. I have now separated the brackets that these statistics are reported in text, to prevent this misunderstanding.

In addition to the articulating the rationale for the between groups design and the reporting issues noted above, some minor issues that also need addressing:

4) Page 3, line 8 AP – cite in full this first time it is mentioned

RESPONSE:

I have now cited MDMA-assisted-psychotherapy (MDMA-AP) in the abstract, where it was first mentioned.

5) Page 5, line 29 – MDMA research has not used an active placebo (though it should!). Please rephrase to increase clarity.

RESPONSE:

Dr Bright is referring to the following line: “All trials of MDMA-AP use an active control group, which receives a placebo drug and the same psychotherapy as the treatment arm”. I have changed this to the following, to ensure it is clear exactly what the control group is: “All trials of MDMA-AP use an active psychotherapy control group, who receive an inactive placebo drug but the same psychotherapy as the treatment arm [12]. The use of an inactive placebo means most participants to become unblinded once they consume MDMA, due to its significant psychoactive effect, thus there is likely to be some uncontrolled placebo effect of MDMA-AP. This has been raised as a serious methodological concern with trials of many repurposed psychoactive compounds [13]… Trial data indicate MDMA-AP is effective compared to active psychotherapy controls, with large
between-groups reductions in CAPS score (range: d=0.91-1.17) [18, 20, 21]. A meta-analysis of phase-II results shows MDMA-AP to be as efficacious as PE [22]. Notably, half of the PE studies used inactive psychotherapy controls, which not only increases between-group differences, but again can also leave participants unblinded to treatment condition in much the same way that an inactive drug control can. Furthermore, none were investigated in treatment-resistant populations. Thus, it seems possible that MDMA-AP may show higher effect sizes than PE, if evaluated in equivalent populations with similar controls.”

7) Page 9, line 13 - I’m not sure what the first sentence of the planned analysis means: “Experimenters were blinded prior to conducting statistical analyses”. perhaps reword for clarity?

RESPONSE:

I have changed the sentence to increase clarity, as follows: “BC de-identified the group condition within the raw data set, so that DW was blinded to treatment condition whilst organising, cleaning and analysing the data.”

Reviewer 2: Dr. Edimansyah Abdin

Comments to the Author:

Thanks for giving opportunity to review this manuscript. This is first study to experimentally assess bias in MH professional attitudes towards the emerging field of psychedelic psychotherapy, particularly MDMA-AP. The study found there is no bias against MDMA-AP, compared to a trial using an unknown, neutrally labelled pharmacotherapy. I have few comments especially in the method section for authors consideration.

8 & 9) I was not clear why randomization based on professions were chosen and why effect size based on minimal clinically difference were chosen when calculating the sample size.

RESPONSE:

We randomized based on treatment condition stratified by profession to maintain internal validity and balance participant numbers in the subgroups compared in the secondary outcomes. A priori we hypothesised that there would be attitudinal differences between the professions and this process (a) minimises the impact of a key confounder on the primary outcome and best allows us to address the question of between profession differences. It is a very commonly used randomisation strategy to equally allocate people with different levels of severity or gender.

We used “minimal clinically significant difference” as the guide for the effect size input into sample size calculations, as this was deemed the lowest effect size we would consider as meaningful when assessing the primary and secondary outcomes. Any sample size calculation is based upon an assumption, ours being that if effect sizes were any lower, they begin to lose any meaningful generalisability.

10) Since the sample size was split into two group and further stratified by professions, I was wondering how authors ensure that the sample size is equal across these 4 strata when data collection was conducted?

RESPONSE:

There was no way to ensure that the sample size was equal between professions, as participants were recruited using online advertisements disseminated through a variety of mental health organisations and the sampling frame for each varied. However, the method of randomising stratified by profession maximised the chances that the sample sizes remained equal between these groups.
11) It is also not clear if the sample size calculation has taken into account whether the power is sufficient for these subgroup. It would be good if authors can provide any reference and justify this approach.

RESPONSE:

The parameters and tool used to calculate the required sample sizes for to assess primary and secondary outcomes is described in the statistics section. Our sample size was sufficient to assess our primary outcomes. Unfortunately, as is almost ubiquitous in trials powered on one main outcome, we did not recruit enough participants to ensure our analyses were powerful enough to detect smaller, but still potentially clinically significant differences – i.e. there was potential for type 2 errors in our secondary outcomes. Regardless, we were still able to detect larger subgroup differences. I have now included an explanation of this limitation in the discussion, as follows:

“We also recruited fewer psychiatrists than psychologists or MH researchers, which limited the power and sensitivity of our secondary and exploratory analyses. There is a chance we missed smaller, but still clinically meaningful differences between our professional subgroups. Despite such issues, we still managed to show differences across professions.”

12) In statistical analysis section, authors should clarify whether normality test had been done before using parametric tests such as ANOVA etc were used.

RESPONSE:

Thank you for your vigilance in pointing out where I forgot to describe the assumptions for our statistical analyses. I have amended this oversight with the following text in the statistics section:

“Assumptions of normality were tested using visual inspection of Q-Q plots and distributions. For two-way ANOVA, assumptions for homogeneity of variances were assessed using Levene’s test. The only variable to violate normality was the strength of objection by treatment arm and thus, we used the Mann-Whitney U test to compare differences between these groups.”

Regarding the normality of distributions, the strength of objection was found to violate the assumption of normality and thus, I have changed the analysis to a non-parametric analysis, using independent samples Mann-Whitney U test.

13) Although randomization was used, the group seems to be different based on age and experience. It would be good if authors can comment on this finding and it might affect their randomization approach and significant findings.

RESPONSE:

Yes, unfortunately despite randomizing across treatment conditions and professions, a baseline difference in age and experience was still found between treatment conditions and professions. It may be that the differences between professions represent an underlying population difference – for example, psychiatrists may actually be older and more experienced than researchers and psychologists. However, these differences, particularly the differences between treatment conditions, are most likely to be random anomalies during the randomisation process.

Dr. Edimansyah Abdin correctly points out that these between group differences may affect our findings. We were also concerned that the subgroup differences we identified were moderated by the differences in age and experience. Given the failure of our randomisation to adequately allocate age and experience equally we also assessed whether these differences may have confounded our results as described in the results section:

“Given the differences in age and experience across professions, we conducted sensitivity analysis using ANCOVAs controlling for age, experience, stance on decriminalisation, and gender. There were no significant covariates, and in all cases a significant interaction of profession and treatment arm remained (Supplementary Figure 2).”
## GENERAL COMMENTS

The authors have made a concerted effort to address the concerns of the reviewers. I believe the manuscript should now be accepted for publication.