**NIH Quality Assessment Tool for Before-After (Pre-Post) Studies With No Control Group**

*Note: the tool has not been standardised and independently published. Researchers have to determine their own parameters for making judgements.*

*CD, cannot determine; NA, not applicable; NR, not reported*

**Eisenstadt, 2021**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Criteria** | **Yes** | **No** | **CD, NA, NR** | **Argumentation/evidence** |
| 1.Was the study question or objective clearly stated? | X |  |  | “Therefore, in this evaluation, we aimed to describe the acceptability, engagement, and preliminary outcomes of using an app (Paradym) designed to promote emotional well-being by adopting novel approaches”. |
| 2.Were eligibility/selection criteria for the study population prespecified and clearly described? |  | X |  | Even though eligibility was clearly described in paper it was not mentioned in pre-registration.  “To be eligible for the user testing, participants needed to be over the age of 18 years and not have a diagnosable mental health condition.” |
| 3.Were the participants in the study representative of those who would be eligible for the test/service/intervention in the general or clinical population of interest? | X |  |  | “Real-world (ie, in-the-wild) data were collected from an international pool of potential users.” |
| 4.Were all eligible participants that met the prespecified entry criteria enrolled? | X |  | CD | No prespecified entry criteria.  Standardised automated links were provided if people met the entry criteria (see figure 1):  “Participants were enrolled in the user testing through a link that was provided on the web advertisement and on the Paradym website. The link provided access to web-based questionnaires for the baseline assessment and instructions on how to download the app from the Apple app store for iPhone users or Google Play for Android users. An automated email was sent to participants 2 weeks after completing the first battery of questionnaires for participants to complete the follow-up questionnaires”. |
| 5.Was the sample size sufficiently large to provide confidence in the findings? |  |  | CD | No information provided on power calculation and minimum sample size needed.  292 expressed interest, 115 completed baseline questionnaires, 34 completed post-questionnaires |
| 6.Was the test/service/intervention clearly described and delivered consistently across the study population? | X |  |  | The intervention was clearly described (see p.4) and delivered consistently – fully automated – across study participants. |
| 7.Were the outcome measures prespecified, clearly defined, valid, reliable, and assessed consistently across all study participants? | X |  |  | Pre-registration available and measure was listed:  “The SWLS is a short 5-item scale designed to measure the global cognitive assessment of satisfaction with one’s life. The estimated time for completion of SWLS has been reported to be approximately 1 minute. The SWLS has been shown to have very high construct validity, with Cronbach α=.85-.87 [59] and moderately high reliability (Cronbach α=.78) [60]. A high score obtained from the SWLS indicates a high level of life satisfaction”. |
| 8.Were the people assessing the outcomes blinded to the participants' exposures/interventions? |  | X |  | Participants were aware they received the intervention and therefore weren’t blinded. |
| 9.Was the loss to follow-up after baseline 20% or less? Were those lost to follow-up accounted for in the analysis? |  | X |  | There was a high level (70.4%) of dropout and this was not accounted for in the analysis:  115 completed baseline questionnaires, 34 completed post-questionnaires  “On the basis of the available data from the participants who completed the evaluation … Paired samples, 2-tailed t tests were conducted on continuous data, and Chi-square tests were used for categorical variables in SPSS”. |
| 10.Did the statistical methods examine changes in outcome measures from before to after the intervention? Were statistical tests done that provided p values for the pre-to-post changes? | X |  |  | See table 3 for p-value and the following:  “descriptive analysis showed that mean scores slightly improved on the SWLS (+1.56), the WHO-5 (+2.03), and the ESAS (+1.75). However, the increase was statistically significant only for the WHO-5 score (t33=2.87)”. |
| 11.Were outcome measures of interest taken multiple times before the intervention and multiple times after the intervention (i.e., did they use an interrupted time-series design)? |  | X | NA |  |
| 12.If the intervention was conducted at a group level (e.g., a whole hospital, a community, etc.) did the statistical analysis take into account the use of individual-level data to determine effects at the group level? |  |  | NA |  |

Additional relevant parameters that could influence the Risk of Bias/study quality based upon researchers’ own judgement:

|  |  |
| --- | --- |
| Evidence: | Impact upon quality/risk of bias of study: |
|  |  |
|  |  |
|  |  |
|  |  |

|  |  |
| --- | --- |
| Overall quality rating: | High Risk of Bias – Low quality |
| Reasoning: | Pre-registration incomplete, no predetermined sample size, high attrition which was not accounted for in analysis, no blinding of participants and outcome assessors. |