Revised Cochrane risk-of-bias tool for randomized trials (RoB 2)

TEMPLATE FOR COMPLETION

Edited by Julian PT Higgins, Jelena Savović, Matthew J Page, Jonathan AC Sterne  
on behalf of the RoB2 Development Group

**Version of 22 August 2019**

The development of the RoB 2 tool was supported by the MRC Network of Hubs for Trials Methodology Research (MR/L004933/2- N61), with the support of the host MRC ConDuCT-II Hub (Collaboration and innovation for Difficult and Complex randomised controlled Trials In Invasive procedures - MR/K025643/1), by MRC research grant MR/M025209/1, and by a grant from The Cochrane Collaboration.



This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](http://creativecommons.org/licenses/by-nc-nd/4.0/).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Study details**   |  |  | | --- | --- | | **Reference** | Bakker, D., Kazantzis, N., Rickwood, D., & Rickard, N. (2018). A randomized controlled trial of three smartphone apps for enhancing public mental health. *Behaviour Research and Therapy, 109,* 75-83. https://doi.org/10.1016/j.brat.2018.08.003 |   **Study design**   |  |  | | --- | --- | | X | Individually-randomized parallel-group trial | | £ | Cluster-randomized parallel-group trial | | £ | Individually randomized cross-over (or other matched) trial |   **For the purposes of this assessment, the interventions being compared are defined as**   |  |  |  |  | | --- | --- | --- | --- | | Experimental: | MoodPrism & MoodKit | Comparator: | Waitlist Control |  |  |  | | --- | --- | | **Specify which outcome is being assessed for risk of bias** | Mental well-being (WEMWBS) |  |  |  | | --- | --- | | **Specify the numerical result being assessed.** In case of multiple alternative analyses being presented, specify the numeric result (e.g. RR = 1.52 (95% CI 0.83 to 2.77) and/or a reference (e.g. to a table, figure or paragraph) that uniquely defines the result being assessed. | Table 2, journal article, Intention To Treat (ITT) mental well-being outcome at pre and post |   **Is the review team’s aim for this result…?**   |  |  | | --- | --- | | £ | to assess the effect of *assignment to intervention* (the ‘intention-to-treat’ effect) | | X | to assess the effect of *adhering to intervention* (the ‘per-protocol’ effect) |   **If the aim is to assess the effect of *adhering to intervention***, select the deviations from intended intervention that should be addressed (at least one must be checked):  X occurrence of non-protocol interventions  £ failures in implementing the intervention that could have affected the outcome  X non-adherence to their assigned intervention by trial participants  **Which of the following sources were obtained to help inform the risk-of-bias assessment? (tick as many as apply)**  X Journal article(s) with results of the trial  £ Trial protocol  £ Statistical analysis plan (SAP)  £ Non-commercial trial registry record (e.g. ClinicalTrials.gov record)  £ Company-owned trial registry record (e.g. GSK Clinical Study Register record)  £ “Grey literature” (e.g. unpublished thesis)  £ Conference abstract(s) about the trial  £ Regulatory document (e.g. Clinical Study Report, Drug Approval Package)  £ Research ethics application  £ Grant database summary (e.g. NIH RePORTER or Research Councils UK Gateway to Research)  £ Personal communication with trialist  £ Personal communication with the sponsor |

## Risk of bias assessment

Responses underlined in green are potential markers for low risk of bias, and responses in red are potential markers for a risk of bias. Where questions relate only to sign posts to other questions, no formatting is used.

**Domain 1: Risk of bias arising from the randomization process**

|  |  |  |
| --- | --- | --- |
| **Signalling questions** | **Comments** | **Response options** |
| **1.1 Was the allocation sequence random?** | “Participants were randomized to one of the four conditions and sent relevant instructions … After participant emails were collected, allocation to condition was performed at random using a repeating sequence on a spreadsheet. Participants were not informed of the other possible condition assignments. Participants were emailed a link to an online Qualtrics survey which administered the initial assessment. The assessment contained their allocated condition's relevant measures and instructions; for example, how to download and set up the allocated app. As MoodKit was a paid app, participants were provided a free download code to redeem on the iTunes Store. Once participants completed all the steps outlined, they provided their email address via a separate online form to indicate completion and maintain participants' anonymity. Participants were emailed a link to the final Qualtrics assessment 30 days later” (Journal article, p.78). | Y  Answer ‘Yes’ if a random component was used in the sequence generation process. Examples include computer-generated random numbers.  Answer ‘No’ if no random element was used in generating the allocation sequence or the sequence is predictable.  Answer ‘No information’ if the only information about randomization methods is a statement that the study is randomized. |
| **1.2 Was the allocation sequence concealed until participants were enrolled and assigned to interventions?** | Y  Answer ‘Yes’ if the trial used any form of remote or centrally administered method to allocate interventions to participants, where the process of allocation is controlled by internet-based randomization service providers.  Answer ‘No’ if there is reason to suspect that the enrolling investigator or the participant had knowledge of the forthcoming allocation. |
| **1.3 Did baseline differences between intervention groups suggest a problem with the randomization process?** | “A univariate ANOVA revealed no significant differences between groups on Age, and a Chi-squared test revealed no significant differences for Gender (all p > .05)” (p. 79).  However, visual inspection of figure 2 baseline scores of prognostic factors indicate that depression and anxiety scores are significantly higher at baseline in Moodkit and MoodPrism compared to the Waitlist control. Additionally, mental well-being scores appear significantly lower in Moodkit and MoodPrism compared to the Waitlist control. No analysis or information is available on this in-text. | PY  Answer ‘Yes’ if there are imbalances that indicate problems with the randomization process, including:  (1) substantial differences between intervention group sizes, compared with the intended allocation ratio; or  (2) a substantial excess in statistically significant differences in baseline characteristics between intervention groups, beyond that expected by chance; or  (3) imbalance in one or more key prognostic factors, or baseline measures of outcome variables, that is very unlikely to be due to chance and for which the between-group difference is big enough to result in bias in the intervention effect estimate. |
| **Risk-of-bias judgement** |  | Some concerns |
| Optional: What is the predicted direction of bias arising from the randomization process? | As experimental groups have lower mental well-being scores at baseline and therefore more potential for an increase in scores compared to the control group. | Favours experimental |

Domain 2: Risk of bias due to deviations from the intended interventions (*effect of adhering to intervention*)

|  |  |  |
| --- | --- | --- |
| **Signalling questions** | **Comments** | **Response options** |
| **2.1. Were participants aware of their assigned intervention during the trial?** | 1.1 “Participants were not informed of the other possible condition assignments. Participants were emailed a link to an online Qualtrics survey which administered the initial assessment. The assessment contained their allocated condition's relevant measures and instructions; for example, how to download and set up the allocated app.” (p.78).  However, the waitlist control group did not engage in any intervention and therefore weren’t blinded, realising they did not participate in an active intervention. So they were likely aware of this.  2.2 intervention was fully automated digital. | PY |
| **2.2. Were carers and people delivering the interventions aware of participants' assigned intervention during the trial?** | N |
| **2.3. [If applicable:] If Y/PY/NI to 2.1 or 2.2: Were important non-protocol interventions balanced across intervention groups?** | No information regarding any non-protocol interventions. | NI |
| **2.4. [If applicable:] Were there failures in implementing the intervention that could have affected the outcome?** |  | NA |
| **2.5. [If applicable:] Was there non-adherence to the assigned intervention regimen that could have affected participants’ outcomes?** | There was non-adherence, however, this likely did not affect participants’ outcomes:  “As is the case with eHealth research, attrition rates of up to 50–60% were expected (Hochheimer et al., 2016). Missing data from 30-day assessments were replaced via multiple imputation as recommended in guidelines for eHealth research (Blankers, Koeter, & Schippers, 2010). This method was chosen because data were considered missing at random (MAR), as attrition rates were relatively equal across groups that used the same assessment platforms and the likelihood of a participant completing the 30-day assessment was moderated by many random and situational factors, such as overall phone use, conscientiousness, interpretation of the importance of completing the assessment, and environmental distractions. The Markov Chain Monte Carlo algorithm was used with five imputations, and all baseline measures were included as predictors to replace missing data from 30-day assessments.” | PN |
| **2.6. If N/PN/NI to 2.3, or Y/PY/NI to 2.4 or 2.5: Was an appropriate analysis used to estimate the effect of adhering to the intervention?** | “To investigate the potential mediating effects of the secondary measures on the primary outcome measures, a series of mediated regression models were used. These analyses were conducted with the PROCESS plug-in for SPSS (A. F. Hayes, 2013) using procedures detailed in Field (2013) and Hayes and Rockwood (2016), which included bootstrapping using 5000 samples. To quantify change over time for each mediator or outcome variable, baseline scores were used as covariates and final scores as outcome variables. This follows Hayes and Rockwood's (2016) recommendations and avoids “self-selection”, regression to the mean, and other biases found in other techniques, such as the use of difference scores. All Beta (β) statistics reported in the regressions are standardized effect sizes, and confidence intervals were inspected to determine statistical significance.” | PN |
| **Risk-of-bias judgement** |  | High |
| Optional: What is the predicted direction of bias due to deviations from intended interventions? |  | Unpredictable |

Domain 3: Missing outcome data

|  |  |  |
| --- | --- | --- |
| **Signalling questions** | **Comments** | **Response options** |
| **3.1 Were data for this outcome available for all, or nearly all, participants randomized?** | See Figure 1. p.78:  MoodKit -> 39 missing at post-intervention out of 78 randomised  MoodPrism -> 52 missing at post-intervention out of 78 randomised  Waitlist -> 25 missing at post-intervention out of 78 randomised | N  The appropriate study population for an analysis of the intention to treat effect is all randomized participants.“Nearly all” should be interpreted as that the number of participants with missing outcome data is sufficiently small that their outcomes, whatever they were, could have made no important difference to the estimated effect of intervention.For continuous outcomes, availability of data from 95% of the participants will often be sufficient. |
| **3.2 If N/PN/NI to 3.1: Is there evidence that the result was not biased by missing outcome data?** | “… data were considered missing at random (MAR), as attrition rates were relatively equal across groups that used the same assessment platforms and the likelihood of a participant completing the 30-day assessment was moderated by many random and situational factors, such as overall phone use, conscientiousness, interpretation of the importance of completing the assessment, and environmental distractions.” (p.78-79). | Y  Evidence that the result was not biased by missing outcome data may come from: (1) analysis methods that correct for bias; or (2) sensitivity analyses showing that results are little changed under a range of plausible assumptions about the relationship between missingness in the outcome and its true value. However, imputing the outcome variable, either through methods such as ‘last-observation-carried-forward’ or via multiple imputation based only on intervention group, should not be assumed to correct for bias due to missing outcome data. |
| **3.3 If N/PN to 3.2: Could missingness in the outcome depend on its true value?** |  | NA  If loss to follow up, or withdrawal from the study, could be related to participants’ health status, then it is possible thatmissingness in the outcome was influenced by its true value. However, if all missing outcome data occurred for documented reasons that are unrelated to the outcome then the risk of bias due to missing outcome data will be low (for example, failure of a measuring device or interruptions to routine data collection).In time-to-event analyses, participants censored during trial follow-up, for example because they withdrew from the study, should be regarded as having missing outcome data, even though some of their follow up is included in the analysis. Note that such participants may be shown as included in analyses in CONSORT flow diagrams. |
| **3.4 If Y/PY/NI to 3.3: Is it likely that missingness in the outcome depended on its true value?** | NA  Reasons for answering ‘Yes’ are:  1.Differences between intervention groups in the proportions of missing outcome data.  2.Reported reasons for missing outcome data provide evidence that missingness in the outcome depends on its true value.  3.Reported reasons for missing outcome data differ between the intervention groups.  4.The circumstances of the trial make it likely that missingness in the outcome depends on its true value. For example, in trials of interventions to treat schizophrenia it is widely understood that continuing symptoms make drop out more likely.  Answer ‘No’ if the analysis accounted for participant characteristics that are likely to explain the relationship between missingness in the outcome and its true value. |
| **Risk-of-bias judgement** | Could potentially change to ‘some concerns’ due to high level of dropout in some groups compared to other groups. However, authors of study have judged this to be at random with relevant analysis so ‘low’ would also be appropriate judgement and will be used for current review. | Low |
| Optional: What is the predicted direction of bias due to missing outcome data? |  | NA |

Domain 4: Risk of bias in measurement of the outcome

|  |  |  |
| --- | --- | --- |
| **Signalling questions** | **Comments** | **Response options** |
| **4.1 Was the method of measuring the outcome inappropriate?** | Standardised mental well-being measure with good validity and reliability: “The Warwick-Edinburgh Mental Well-being Scale (WEMWBS; Stewart-Brown & Janmohamed, 2008) is a 14-item assessment of mental wellbeing. The frequency of positive psychological experiences over the past two weeks (e.g. “been feeling close to other people”) are rated on a five-point scale from 1 (none of the time) to 5 (all of the time). The WEMWBS shares high correlations with measures of life satisfaction and other measures of well-being, and has high internal reliability, Cronbach's α = 0.91 (Tennant et al., 2007). In this study, Cronbach's α = 0.92.” (p.77-78). | N  Answer ‘Yes’ or ‘Probably yes’ if the method of measuring the outcome is inappropriate, for example because:  (1) it is unlikely to be sensitive to plausible intervention effects (e.g. important ranges of outcome values fall outside levels that are detectable using the measurement method); or  (2) the measurement instrument has been demonstrated to have poor validity. |
| **4.2 Could measurement or ascertainment of the outcome have differed between intervention groups?** | Digitally obtained and standardised mental well-being measure: “Participants were emailed a link to an online Qualtrics survey which administered the initial assessment. The assessment contained their allocated condition's relevant measures and instructions. … Participants were emailed a link to the final Qualtrics assessment 30 days later. This assessment contained the final measure.” (p.78). | N  Comparable methods of outcome measurement (data collection) involve the same measurement methods and thresholds, used at comparable time points. Differences between intervention groups may arise because of ‘diagnostic detection bias’ in the context of passive collection of outcome data, or if an intervention involves additional visits to a healthcare provider, leading to additional opportunities for outcome events to be identified. |
| **4.3 If N/PN/NI to 4.1 and 4.2: Were outcome assessors aware of the intervention received by study participants?** | Outcome assessors (participants) were blinded to the intervention received: “Participants were not informed of the other possible condition assignments.”  Despite not informing them, participants will have likely been aware whether they received an active intervention or waitlist control. | PY  Answer ‘No’ if outcome assessors were blinded to intervention status. For participant-reported outcomes, the outcome assessor is the study participant. |
| **4.4 If Y/PY/NI to 4.3: Could assessment of the outcome have been influenced by knowledge of intervention received?** |  | PY  Knowledge of the assigned intervention could influence participant-reported outcomes (such as level of pain), observer-reported outcomes involving some judgement, and intervention provider decision outcomes. They are unlikely to influence observer-reported outcomes that do not involve judgement, for example all-cause mortality. |
| **4.5 If Y/PY/NI to 4.4:** **Is it likely that assessment of the outcome was influenced by knowledge of intervention received?** | PN  This question distinguishes between situations in which (i) knowledge of intervention status could have influenced outcome assessment but there is no reason to believe that it did (assessed as ‘Some concerns’) from those in which (ii) knowledge of intervention status was likely to influence outcome assessment (assessed as ‘High’). When there are strong levels of belief in either beneficial or harmful effects of the intervention, it is more likely that the outcome was influenced by knowledge of the intervention received. Examples may include patient-reported symptoms in trials of homeopathy, or assessments of recovery of function by a physiotherapist who delivered the intervention. |
| **Risk-of-bias judgement** |  | Some concerns |
| Optional: What is the predicted direction of bias in measurement of the outcome? |  | Unpredictable |

Domain 5: Risk of bias in selection of the reported result

|  |  |  |
| --- | --- | --- |
| **Signalling questions** | **Comments** | **Response options** |
| **5.1 Were the data that produced this result analysed in accordance with a pre-specified analysis plan that was finalized before unblinded outcome data were available for analysis?** | There is no availability of a pre-specified analysis plan or pre-registration which preceded the reported results in the journal article. | NI  If the researchers’ pre-specified intentions are available in sufficient detail, then planned outcome measurements and analyses can be compared with those presented in the published report(s). To avoid the possibility of selection of the reported result, finalization of the analysis intentions must precede availability of unblinded outcome data to the trial investigators. |
| **Is the numerical result being assessed likely to have been selected, on the basis of the results, from...** |  |  |
| **5.2. ... multiple eligible outcome measurements (e.g. scales, definitions, time points) within the outcome domain?** | Pre and post outcomes are available and there’s not many other ways in which the mental well-being outcome could’ve been measured differently other than using a different tool entirely. | PN  Answer ‘Yes’ or ‘Probably yes’ if:  There is clear evidence (usually through examination of a trial protocol or statistical analysis plan) that a domain was measured in multiple eligible ways, but data for only one or a subset of measures is fully reported (without justification), and the fully reported result is likely to have been selected on the basis of the results. Selection on the basis of the results can arise from a desire for findings to be newsworthy, sufficiently noteworthy to merit publication, or to confirm a prior hypothesis. For example, trialists who have a preconception, or vested interest in showing, that an experimental intervention is beneficial may be inclined to report outcome measurements selectively that are favourable to the experimental intervention.  Answer ‘No’ or ‘Probably no’ if:  There is clear evidence (usually through examination of a trial protocol or statistical analysis plan) that all eligible reported results for the outcome domain correspond to all intended outcome measurements.  Or  There is only one possible way in which the outcome domain can be measured (hence there is no opportunity to select from multiple measures).  or  Outcome measurements are inconsistent across different reports on the same trial, but the trialists have provided the reason for the inconsistency and it is not related to the nature of the results.  Answer ‘No information’ if:  Analysis intentions are not available, or the analysis intentions are not reported in sufficient detail to enable an assessment. |
| **5.3 ... multiple eligible analyses of the data?** | All eligible results (group x time at pre-post for all groups) seem to have been reported. | PN  There is clear evidence (usually through examination of a trial protocol or statistical analysis plan) that a measurement was analysed in multiple eligible ways, but data for only one or a subset of analyses is fully reported (without justification), and the fully reported result is likely to have been selected on the basis of the results. |
| **Risk-of-bias judgement** | There’s no pre-specified analysis plan available prior to the journal article. | Some concerns |
| Optional: What is the predicted direction of bias due to selection of the reported result? |  | Unpredictable |

Overall risk of bias

|  |  |  |
| --- | --- | --- |
| **Risk-of-bias judgement** | Only minor concerns due to some baseline differences which are considered not to arise at random and missing pre-specified analysis plan. | High |
| Optional: What is the overall predicted direction of bias for this outcome? |  | Favours experimental |



This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](http://creativecommons.org/licenses/by-nc-nd/4.0/).