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** | Champion, L., Economides, M., & Chandler, C. (2018). The efficacy of a brief app-based mindfulness intervention on psychosocial outcomes in healthy adults: A pilot randomised controlled trial. PLoS ONE, 13(12), e0209482. https://doi.org/10.1371/journal.pone.0209482 |   **Study design**   |  |  | | --- | --- | | £ | 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: | ‘Headspace’ mindfulness meditation app | Comparator: | Waitlist control |  |  |  | | --- | --- | | **Specify which outcome is being assessed for risk of bias** | Subjective well-being - Satisfaction With Life Scale (Diener, 1975) |  |  |  | | --- | --- | | **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. | Both completers only and imputed values for experimental and control at baseline, day 10 and day 30 (post-intervention), see table 2, p.10 in journal paper. |   **Is the review team’s aim for this result…?**   |  |  | | --- | --- | | X | to assess the effect of *assignment to intervention* (the ‘intention-to-treat’ effect) | | £ | 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):  £ occurrence of non-protocol interventions  £ failures in implementing the intervention that could have affected the outcome  £ 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)  X 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?** | “The remaining participants were randomised (using simple randomization) via a computer-generated sequence with equal weight to the MM group or WL control group. Sequence generation and randomization was performed by the research team, who were not formally blinded to group allocation”.  Although the research team wasn’t blinded, a remote internet-based randomisation was performed. | 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?** | “Age ranged from 25–59 years (M = 39.13, SD = 5.70, 37 females and 25 males; see Table 1 for characteristics of the sample). Note that whilst mean age was similar between groups, randomisation resulted in a larger proportion of females than males in the wait-list group, likely reflecting a limitation of simple randomisation.  Baseline scores suggest that participants in both groups had average levels of satisfaction with life (c.f. [37]), average to high levels of stress (c.f. [40]), and moderate levels of resilience (c.f. [48]) prior to engaging with the study. Mean GHQ-28 scores at baseline were 3.72 (SD = 3.03) in the MM group and 3.36 (SD = 3.16) in the WL group. These scores are below the general threshold for possible psychological distress [49], but suggest that at least some of the participants may have been candidates for distress. Scores did not differ between groups at baseline on any measures (two-sample t-tests, all p > 0.05).” | Y  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? |  | Unpredictable |

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?** | 2.1 Supplementary materials, study protocol, invitation email:  “Participants will be randomly allocated to either group A or group B. Group A will be asked to listen to Headspace for ideally 10 minutes a day (or every other day) on the tube, at home, or anywhere else for 30 days, and complete a 10-minute questionnaire at the start of the study, and following 10 and 30 days.  Group B will be asked to carry on with their everyday lives as usual, not listening to Headspace, but completing a questionnaire at the start of the study, and again following 10 and 30 days. When group B have completed the final questionnaire at day 30, they will have free access to Headspace for the following 30 days.”  2.2 Intervention is delivered by autonomous digital technology. | Y |
| **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?** | There is no information to regarding whether participants deviated from the intended intervention due to the trial context. | 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?** | Participants in the MM group reported engaging with the app an average of 6.21 times (SD = 2.65) between baseline and day 10 (min = 1, max = 10), with 22 out of 29 participants completing 5 or more out of the 10 sessions. Participants reported engaging with the MM app 11.66 times (SD = 6.16) between day 11 and 30 (min = 1, max = 20), with 17 out of 29 participants completing 10 or more out of the 20 sessions. Six participants completed 25 or more out of a maximum of 30 sessions.  Participants that rated the intervention as being relatively easy between days 10 and 30 were likely to have larger increases in satisfaction with life (ß = 1.264, p = 0.006), larger decreases in stress (ß = 1.471, p = 0.015), and larger increases in resilience (ß = 2.570, p = 0.003) over the same time period. Further, participants with higher GHQ-28 scores at baseline (higher levels of baseline distress) were likely to have larger increases in satisfaction with life (ß = 0.508, p = 0.041) and larger increases in resilience (ß = 1.231, p = 0.009) between days 10 and 30 of the MM intervention. Lastly, older participants were likely to have a larger increase in satisfaction with life between baseline and day 10 (ß = 0.323, p = 0.049), whilst females were likely to experience larger decreases in stress between day 10 and day 30 than males (ß = 4.794, p = 0.019).” | PY |
| **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?** | “The present study included both an intention-to-treat (ITT) and complete-case analysis. For the former, outcome scores at baseline, day 10, and day 30 were analysed using linear mixed effects models, as these have the ability to handle missing data and are considered to be superior to other ITT approaches such as ‘last observation carried forward’ [42,43]. A number of nested models were fit using maximum likelihood and compared using likelihood ratio tests. For all outcomes, the winning model included time (coded as 0, 10 & 30), group (coded as 1 & 0) and their interaction as fixed effects variables, and random intercepts and slopes across participants. Covariance between the random intercept and slope was modelled for outcomes in which this improved the model fit”.  “Exploratory multiple regressions were performed to investigate the influence of demographics, self-rated intervention enjoyment / level of difficulty, and intervention engagement on change in outcome scores from pre to post intervention (in the MM group only).” | PN  (Due to exploratory nature of analysis) |
| **Risk-of-bias judgement** |  | High |
| Optional: What is the predicted direction of bias due to deviations from intended interventions? | Due to failure to adhere to intervention. | Towards null |

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?** | Figure 1. p.8, journal article:  74 randomised, 38 to headspace and 36 to Waitlist control.  9 are missing in the headspace condition and 3 in the waitlist control. | 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?** | “In order to rule out the possibility that attrition could have influenced our estimates of Cohen’s d, we used multiple imputation (MI) to impute missing values and re-calculated Cohen’s d using the full imputed dataset [45]. The MI procedure used the data augmentation method [46], with default values of 10 independent chains of length 100. 95% confidence intervals (CIs) were calculated for all Cohen’s d values using equations 15 and 18 (for within-group effect sizes) and equations 15 and 16 (for between-group effect sizes) from Nakagawa and Cuthill, 2007 [47]. … For each outcome, we used multiple imputation to impute missing values and then calculated a between-group Cohen’s d effect size between baseline and day 30. This revealed a Cohen’s d of 0.57 (95% CI: 0.10 to 1.04) for satisfaction with life, 1.42 (95% CI: 0.90 to 1.94) for perceived stress, and 0.63 (95% CI: 0.16 to 1.10) for resilience (see Table 2).”  Visual inspection of table 2. Indicates very minor differences. | 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** |  | 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?** | “Satisfaction with Life Scale. The Satisfaction with Life Scale (SWLS; [37]) is a 5-item scale that assesses global satisfaction (e.g. ‘so far I have gotten the important things I want in my life’). Individuals indicate their degree of agreement or disagreement with the items using a 7-point Likert scale (1 = strongly disagree to 7 = strongly agree). The scale is shown to have strong psychometric properties [38] including internal consistency (α = .87) and a high testretest reliability (α = .82; [39]).” | 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?** | Standardised digital measures were conducted the same across groups:  “Participants were then emailed a link to the baseline questionnaires (see Measures below), which they were instructed to complete immediately, and provided with a description of what to expect from the study over the following 30 days. …  [MM group] Participants were emailed the study questionnaires again following 10 days, which included a set of user experience questions and a note encouraging them to continue using the app each day until the third and final set of questionnaires, which was sent to participants on day 30. Participants that did not complete the final set of questionnaires were sent up to three reminder emails, after which they were considered to have withdrawn.  The WL group were informed that they would not be engaging with any MM content until day 30 of the study (see appendix of ‘S1 Protocol’ for briefing email). As with the MM group, they were instructed to complete follow-up questionnaires at days 10 and 30, and received up to three follow-up emails if they failed to do so.” | 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?** | Participants (assessors) were aware: “The WL group were informed that they would not be engaging with any MM content until day 30 of the study (see appendix of ‘S1 Protocol’ for briefing email).” | Y  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?** | Self-report by participants (assessors). | 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?** | The study protocol includes a pre-specified analysis plan, information about the aims and objectives as well as outcome measures/timepoints. Although there’s no clear date of this study protocol and registration of the study was done retrospectively, it is assumed this study protocol was used to obtain ethical approval prior to the study and was therefore prospective. | PY  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?** | Protocol includes all relevant measures and timepoints reported in final research paper. | 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?** | Analysis of the data was pre-specified in study protocol and received ethical approval – everything was fully 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** |  | Low |
| Optional: What is the predicted direction of bias due to selection of the reported result? |  | NA |

Overall risk of bias

|  |  |  |
| --- | --- | --- |
| **Risk-of-bias judgement** |  | High |
| Optional: What is the overall predicted direction of bias for this outcome? |  | Towards null |



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