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** | Brazier, A., Larson, E., Xu, Y., Judah, G., Egan, M., Burd, H., & Darzi, A. (2022). 'Dear Doctor': a randomised controlled trial of a text message intervention to reduce burnout in trainee anaesthetists. Anaesthesia, 77(4), 405–415. https://doi.org/10.1111/anae.15643 |   **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: | ‘Dear doctor’ text message intervention | Comparator: | Waitlist control |  |  |  | | --- | --- | | **Specify which outcome is being assessed for risk of bias** | Mental well-being (Short 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. | Pre-post mental well-being scores of experimental intervention and control group. Journal article, table 3, p.411. |   **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)  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?** | Journal article (p.407): “Participants were randomly assigned to intervention or control conditions with a 1:1 allocation ratio using a random number generator (Stata, StataCorp, College Station, TX, USA). Randomisation was stratified by training year and across five broad training regions.” | 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?** | “We calculated z-scores to check the balance of randomisation across the control and intervention groups”  Randomisation balance checks (Supplementary materials, table S5) indicate no differences between groups at baseline. | N  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** |  | Low |
| Optional: What is the predicted direction of bias arising from the randomization process? |  | NA |

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 “Participants were not blinded to their condition assignment: intervention group participants received the intervention; control participants did not.”  2.2 Researchers were aware of the received intervention:” Researchers were not blinded after condition assignment in order to determine how many participants from each condition responded to outcome surveys.”  However, the intervention was delivered fully autonomously digital. | 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?** | Participants appear to have used/been offered other support throughout the trial period, however, there is no evidence that this differed but exposure between groups was balanced: free text comments “These comments highlighted the high availability of other wellbeing resources during COVID-19". | PY |
| **2.4. [If applicable:] Were there failures in implementing the intervention that could have affected the outcome?** | Deviations due to the trail context were limited: “Control group participants were asked if they had seen any intervention messages despite not being assigned to receive them; for example colleagues may have shared the messages, to determine intervention spill over (response options were ‘I didn’t see any’, ‘I saw some messages (e.g. 1–4 messages approximately)’, ‘I saw many messages (e.g. 5 or more)’). … Almost all intervention text messages were successfully delivered (99.6%). Spill over between groups was limited: 74 (94%) control group participants reported that they did not see any messages and the remaining 5 (6%) reported seeing some, for example, ‘1–4 messages’.” | NA |
| **2.5. [If applicable:] Was there non-adherence to the assigned intervention regimen that could have affected participants’ outcomes?** | “Female participants (v2 = 8.07, p = 0.018), those with lower burnout (t = -2.42, p = 0.016), higher ‘meaningful’ score (t = 3.29, p = 0.001) and ‘valued’ score (t = 2.97, p = 0.003) were more likely to sign up for the trial after completing the baseline survey. In total, 153 participants (55% of those randomised) completed the final survey. This response rate was balanced between the control (79/153) and intervention groups (74/153); see Figure 1 for recruitment diagram. … Despite the high attrition, there were no significant differences in participant characteristics (v2 s < 3.80, ps > 0.15) or baseline outcomes measures (ts < 0.81, ps > 0.42) between those who did and did not complete the final survey (see also online Supporting Information Table S4). Randomisation was wellbalanced in the final sample: participants in the two trial groups did not differ in terms of profile or covariates (see online Supporting Information Table S5)”. | 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?** | We used OLS regression to test the effect of the intervention on the primary outcome measures (from both the interim and final surveys). Effects were adjusted for sex, training year, training location and baseline measurement. | 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?** | Figure 1. p.410:  279 randomised, 140 to control condition and 139 to intervention.  “274 trainees (18% of the original cohort) consented to participate and remained in the trial throughout the trial period. … In total, 153 participants (55% of those randomised) completed the final survey.”  Thus, 45% dropout – total 121 participants. | 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?** | “This response rate was balanced between the control (79/153) and intervention groups (74/153); see Figure 1 for recruitment diagram. Despite the high attrition, there were no significant differences in participant characteristics (v2 s < 3.80, ps > 0.15) or baseline outcomes measures (ts < 0.81, ps > 0.42) between those who did and did not complete the final survey (see also online Supporting Information Table S4). Randomisation was wellbalanced in the final sample: participants in the two trial groups did not differ in terms of profile or covariates (see online Supporting Information Table S5).” | PY  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?** | “Female participants (v2 = 8.07, p = 0.018), those with lower burnout (t = -2.42, p = 0.016), higher ‘meaningful’ score (t = 3.29, p = 0.001) and ‘valued’ score (t = 2.97, p = 0.003) were more likely to sign up for the trial after completing the baseline survey. In total, 153 participants (55% of those randomised) completed the final survey. This response rate was balanced between the control (79/153) and intervention groups (74/153); see Figure 1 for recruitment diagram. … Despite the high attrition, there were no significant differences in participant characteristics (v2 s < 3.80, ps > 0.15) or baseline outcomes measures (ts < 0.81, ps > 0.42) between those who did and did not complete the final survey (see also online Supporting Information Table S4). Randomisation was wellbalanced in the final sample: participants in the two trial groups did not differ in terms of profile or covariates (see online Supporting Information Table S5)”. | 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?** | Well-being was measured “using the Short Warwick-Edinburgh Mental Well-being scale [22].”  22. Stewart-Brown S, Tennant A, Tennant R, Platt S, Parkinson J, Weich S. Internal construct validity of the Warwick-Edinburgh mental well-being scale (WEMWBS): a Rasch analysis using data from the Scottish health education population survey. Health and Quality of Life Outcomes 2009; 7: 15. | 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?** | Although it should be recognised that interim measures did differ between groups (see below). The result of interest for the current RoB assessment is baseline and final outcomes, which did not differ between groups.  “Outcome measures were self-reported via online surveys, described as well-being surveys to minimise the risk of sensitising respondents to burnout. We used the interim survey to detect if there was a backfire effect, predetermined as a worsening by one standard deviation in either of the primary outcomes. In the event of identifying a backfire effect, we planned to pause the intervention and discuss its adaptation or cancellation with the study Advisory Board. We incentivised completion of the interim and final surveys through a charitable donation of £3 to lifebox.org for each completed survey.” | 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 were not blinded to their condition assignment: intervention group participants received the intervention; control participants did not.” | 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?** | It was a self-report measure and participants were aware of the received intervention. | Y  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?** | Pre-registration is available including information about study design and measures, however, no pre-specified information is available on the analysis of data. This is only available in journal article which doesn’t precede study findings. | N  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-registration document:  **“Primary outcome measure**  1. Wellbeing is measured using the Short Warwick-Edinburgh Mental Wellbeing Scale [SWEMWBS] (7 items) at baseline, 3 months and at the end of the intervention.” | N  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?** | Pre-registration document:  **“Primary outcome measure**  1. Wellbeing is measured using the Short Warwick-Edinburgh Mental Wellbeing Scale [SWEMWBS] (7 items) at baseline, 3 months and at the end of the intervention.” | N  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** |  | 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** |  | High |
| Optional: What is the overall predicted direction of bias for this outcome? |  | Unpredictable |



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