

Adapted Downs and Black checklist (Downs and Black, 1998)

Reporting (score range: 0 to 8)	
1. Is the hypothesis/aim/objective of the study clearly described?	Yes = 1 No = 0
2. Are the main outcomes to be measured clearly described in the Introduction or Methods section? If the main outcomes are first mentioned in the Results section, the question should be answered no.	Yes = 1 No = 0
3. Are the characteristics of the patients included in the study clearly described? In cohort studies and trials, inclusion and/or exclusion criteria should be given. In case-control studies, a case-definition and the source for controls should be given.	Yes = 1 No = 0
4. Are the interventions of interest clearly described? Treatments and placebo (where relevant) that are to be compared should be clearly described.	Yes = 1 No = 0
6. Are the main findings of the study clearly described? Simple outcome data (including denominators and numerators) should be reported for all major findings so that the reader can check the major analyses and conclusions. (This question does not cover statistical tests which are considered below).	Yes = 1 No = 0
7. Does the study provide estimates of the random variability in the data for the main outcomes? In non-normally distributed data, the inter-quartile range of results should be reported. In normally distributed data the standard error, standard deviation or confidence intervals should be reported. If the distribution of the data is not described, it must be assumed that the estimates used were appropriate and the question should be answered yes.	Yes = 1 No = 0
9. Have the characteristics of patients lost to follow-up been described? This should be answered yes where there were no losses to follow-up or where losses to follow-up were so small that findings would be unaffected by their inclusion. This should be answered no where a study does not report the number of patients lost to follow-up.	Yes = 1 No = 0
10. Have actual probability values been reported? (e.g., 0.035 rather than <0.05) for the main outcomes except where the probability value is less than 0.001?	Yes = 1 No = 0
External validity (score range: 0 to 2)	
11. Were the subjects asked to participate in the study representative of the entire population from which they were recruited? The study must identify the source population for patients and describe how the patients were selected. Patients would be representative if they comprised the entire source population, an unselected sample of consecutive patients, or a random sample. Random sampling is only feasible where a list of all members of the relevant population exists. Where a study does not report the proportion of the source population from which the patients are derived, the question should be answered as unable to determine.	Yes = 1 No = 0 Unable to determine = 0
12. Were those subjects who were prepared to participate representative of the entire population from which they were recruited? The proportion of those asked who agreed should be stated. Validation that the sample was representative would include demonstrating that the distribution of the main confounding factors was the same in the study sample and the source population.	Yes = 1 No = 0 Unable to determine = 0

Internal validity (score range: 0 to 4)	
16. If any of the results of the study were based on “data dredging”, was this made clear? Any analyses that had not been planned at the outset of the study should be clearly indicated. If no retrospective unplanned subgroup analyses were reported, then answer yes.	Yes = 1 No = 0 unable to determine = 0
17. In trials and cohort studies, do the analyses adjust for different lengths of follow-up of patients, or in case-control studies, is the time period between the intervention and outcome the same for cases and controls? Where follow-up was the same for all study patients the answer should yes. If different lengths of follow-up were adjusted for by, for example, survival analysis the answer should be yes. Studies where differences in follow-up are ignored should be answered no.	Yes = 1 No = 0 unable to determine = 0
18. Were the statistical tests used to assess the main outcomes appropriate? The statistical techniques used must be appropriate to the data. For example, nonparametric methods should be used for small sample sizes. Where little statistical analysis has been undertaken but where there is no evidence of bias, the question should be answered yes. If the distribution of the data (normal or not) is not described, it must be assumed that the estimates used were appropriate and the question should be answered yes.	Yes = 1 No = 0 unable to determine = 0
20. Were the main outcome measures used accurate (valid and reliable)? For studies where the outcome measures are clearly described, the question should be answered yes. For studies which refer to other work or that demonstrates the outcome measures are accurate, the question should be answered as yes.	Yes = 1 No = 0 unable to determine = 0
Power (score range: 0 to 1)	
27. Did the study mention having conducted a power analysis to determine the sample size needed to detect a significant difference in effect size for one or more outcome measures?	Yes = 1 No = 0

Risk Of Bias In Non-randomized Studies of Interventions tool (ROBINS-I; Sterne et al., 2016)

Recruitment / Selection Bias (score range: 0 to 4)	
1. Were key covariates taken into account? *ROBINS-I domain 1: Bias due to confounding*	0 = No mention of adjustment for covariates or do not specify which covariates were adjusted for 1 = Adjustment for specified covariates but the covariates were not selected <i>a priori</i> 2 = Appropriate covariates were selected <i>a priori</i> and adjusted for in analyses (or a randomised crossover/control trial)
2. How were comorbid health/medical conditions treated? *ROBINS-I domain 1: Bias due to confounding* *ROBINS-I domain 2: Bias in selection of participants into the study*	0 = Comorbid health/medical conditions were not assessed or reported 1 = Comorbid health/medical conditions were adjusted for as covariates in analyses 2 = Participants with comorbid health/medical conditions were removed or analysed separately
3. Are the individuals selected to participate in the study likely to be representative of the target population? *ROBINS-I domain 2: Bias in selection of participants into the study*	0 = Non-random selection process, or sampling not reported 1 = Entire sample was randomly selected from target population
4. Was an <i>a priori</i> sample size calculation used to indicate adequate statistical power for detecting main effects? *ROBINS-I domain 2: Bias in selection of participants into the study*	0 = No 1 = Yes
Measurement Precision (score range: 0 to 12)	
5. Was there an adequate baseline period prior to the stress task? *ROBINS-I domain 3: Bias in classification of interventions*	0 = Duration of baseline period not reported or inappropriate 1 = Appropriate baseline period prior to the stress task
6. Measurement of physical activity/sedentary behavior *ROBINS-I domain 3: Bias in classification of interventions*	0 = Measured via a self-report tool 1 = Measured via a wearable device 2 = Measured via a combination of a self-report tool and a wearable device
7. Duration of physical activity/sedentary behavior measurement *ROBINS-I domain 3: Bias in classification of interventions*	0 = Measured for less than 7 days 1 = Measured for 7 days or more
8. Was there a stress protocol manipulation check? *ROBINS-I domain 3: Bias in classification of interventions*	0 = None reported 1 = Yes, using a self-report measure only 2 = Yes, with both self-report and objective measures of stress

<p>9. Consistency in the stress reactivity protocol</p> <p>*ROBINS-I domain 4: Bias due to deviations from intended interventions*</p>	<p>0 = Low consistency (e.g., stress protocol conducted at different times of the day, conducted in different laboratories, conducted by multiple different researchers, deviations from protocol mentioned etc.)</p> <p>1 = High consistency (e.g., stress protocol conducted at the same time of the day, conducted in the same highly controlled laboratory environment, conducted by the same researcher, no mention of deviation from protocol mentioned etc.)</p>
<p>10. Consistency in the measurement of physical activity/sedentary behaviour</p> <p>*ROBINS-I domain 4: Bias due to deviations from intended interventions*</p>	<p>0 = No consistency in measurement of physical activity/sedentary behaviour across participants (e.g., different questionnaires/device used, or wear time not reported for device-based techniques)</p> <p>1 = Same physical activity/sedentary behaviour questionnaire/device used for all participants, adequate wear time reported</p>
<p>11. Equipment used to measure psychobiological stress reactivity</p> <p>*ROBINS-I domain 6: Bias in measurement of outcomes*</p>	<p>0 = non-validated measure of psychobiological stress reactivity or measurement technique not reported</p> <p>1 = Validated measure of psychobiological stress reactivity (e.g., validated blood pressure cuff and blood pressure measurement protocol)</p>
<p>12. Accounting for baseline when calculating stress reactivity</p> <p>*ROBINS-I domain 6: Bias in measurement of outcomes*</p>	<p>0 = Baseline not accounted for</p> <p>1 = Baseline accounted for (e.g., baseline physiology included as covariate, reactivity calculated as relative change, or analyses include time as within-subject factor)</p>
<p>13. Appropriate adherence criteria before the stress reactivity protocol</p> <p>*ROBINS-I domain 6: Bias in measurement of outcomes*</p>	<p>0 = Adherence criteria (e.g., not consuming stimulants at least 4 hours prior to testing) not reported</p> <p>1 = Adherence criteria reported, but there was not an adequate time frame or there were missing key adherence criterion</p> <p>2 = Adherence criteria was reported with adequate time frame (e.g., no coffee 4 hours prior) and all key criterion included (e.g., no stimulants, exercise, or food).</p>
<p>Results / Discussion (score range: 0 to 3)</p>	
<p>14. What is the quality of the inferences based on the data?</p> <p>*ROBINS-I domain 7: Bias in selection of the reported result*</p>	<p>0 = Poor (did not report on hypothesized outcome and/or incorrect interpretation)</p> <p>1 = Fair (e.g., interpreting marginal results as significant, overstating the robustness of results, only reporting certain findings.</p> <p>2 = Good (interpretations are based on the findings, reported all findings).</p>
<p>15. Rates of missing data</p> <p>*ROBINS-I domain 5: Bias due to missing data*</p>	<p>0 = Large amounts of missing data, inappropriate approaches to handle missing data</p> <p>1 = Small amounts of missing data, appropriate approaches taken to handle missing data</p>