

Table 3. Methodological quality assessment of female athletic performance studies using the modified Downs and Black¹⁸ checklist

Key: UTD: Unable to Determine (Scored as 0); Q: Question; Mod: Moderate

Author	Cabre et al. ⁹	Carmichael et al. ¹⁰	Osborne et al. ³⁴	Peltonen et al. ³⁸	Smith et al. ⁴²	Thompson et al. ⁴⁹
Reporting						
Is the hypothesis/aim/objective of the study clearly described? Yes = 1 No = 0	1	1	1	1	1	1
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, answer no. Yes = 1 No = 0	1	1	1	1	1	1
Are the characteristics of the participants included in the study clearly described? In observational studies, inclusion and/or exclusion criteria should be given. In case-control studies, inclusion and/or exclusion and the source of controls should be given. Yes = 1 No = 0	1	1	1	1	1	1
Were the tested menstrual cycle phases clearly described? Answer yes if the precise criteria used to define phase were provided, answer no if the exact phase tested cannot be ascertained (e.g., vague language such as “early” or “late” were used, without defining the criteria) Yes = 1 No = 0	1	0	1	1	1	1
Are the main findings of the study clearly described? Simple outcome data should be reported for all major findings so the reader can check the major analyses and conclusions. This does not cover statistical tests which are addressed in other questions. Yes = 1 No = 0	1	1	1	1	1	1
Does the study provide estimates of the random variability in the data for the main outcomes? In non-normal data, inter-quartile range should be reported. In normal data, standard deviation, standard error or confidence intervals should be reported. Yes = 1 No = 0	1	0	1	1	1	0
External validity						
Were the subjects confirmed as non-hormonal contraceptive users, for at least three months prior to participation? Yes = 1 No = 0 Unable to determine = 0	1	1	1	1	1	1
Internal validity – bias						
Was at least one familiarization trial conducted prior to exercise testing? Yes = 1 No = 0 Unable to determine = 0	0 (UTD)	1	1	1	1	1
Were the exercise test conditions adequately standardised (taking into consideration factors including time of day, prior nutritional intake [including caffeine] and prior exercise). Yes (all relevant factors standardised) = 2 Yes (some relevant factors standardised) = 1 Exercise testing unstandardized = 0 Unable to determine = 0	1	0 (UTD)	1	1	1	1
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 should be clearly indicated. If no retrospective subgroup analyses were reported, then answer yes. Yes = 1 No = 0 Unable to determine = 0	0 (UTD)	0 (UTD)	0 (UTD)	0 (UTD)	0 (UTD)	0 (UTD)
Were statistical tests used to assess the main outcomes appropriate? The statistical techniques used must be appropriate to the data and the research question. Yes = 1 No = 0 Unable to determine = 0	1	1	1	1	1	1
Were the main outcome measures used accurate (i.e., valid and reproducible)? For studies where the validity and reproducibility of outcome measures are clearly described, the question should be answered yes. For studies which refer to other work that demonstrates the outcome measures are accurate, answer yes. Yes = 1 No = 0 Unable to determine = 0	1	1	1	1	1	1
Internal validity – confounding (selection bias)						
Was the order of phase testing randomised? Yes = 1 No = 0 Unable to determine = 0	1	0 (UTD)	0 (UTD)	1	1	0 (UTD)
Power						
Did the study have sufficient power to detect an <i>a priori</i> specified scientifically important effect at a pre-determined probability threshold? Answer yes if they included a power calculation, and no if not. Yes = 1 No = 0	0	0	0	0	0	0
Was study retention > 85%? Yes = 1 No = 0 Unable to determine = 0	1	1	1	0 (UTD)	0	0
GRADE (assign an <i>a-priori</i> study quality rating based on the modified Downs and Black checklist, so all studies will start out as being of “high”, “moderate”, “low”, “very low”).						
Identify if menstrual cycle phase was confirmed using blood samples. If yes, the <i>a priori</i> rating is maintained, and this is the final study quality rating. If not, the study is downgraded a level (e.g., a study that started out as high, drops to moderate).	No	No	Yes	Yes	Yes	Yes
Identify if menstrual cycle phase was confirmed using ovulation kits. If yes, the Q1. rating is maintained. If no, the study is downgraded another level (e.g., a study that started out high, drops to low). This means that the maximum rating that any study that does not use blood analysis or ovulation kits is “low” or “very low”.	Yes	No	Yes	Yes	Yes	Yes
Rating	12	9	12	12	12	10
Downs and Black checklist (maximum score attainable = 16). Study quality was categorised as follows: “high”: 14 – 16; “moderate”: 10 – 13; “low”: 6 – 9; “very low”: 0 – 5).						
A priori	Mod	Low	Mod	Mod	Mod	Mod
Final classification	Low	Very Low	Mod	Mod	Mod	Mod