**Supplementary Figure S1. Critical Appraisal**

We assessed study limitations of primary research studies using five dimensions:
- Selection bias and confounding
  - Selection bias considered systematic differences between baseline characteristics of the groups that are being compared. The risk is low in randomized controlled trials where the trial investigator randomly assigns participants to the intervention and control group (assuming that the random sequence was correctly generated and allocation concealment was maintained). Most prone to bias are observational studies where participants self-select the intervention or exposure because the compared groups may differ in other characteristics even before the intervention was introduced. These characteristics or confounders are likely to influence any observed difference between the intervention and control group, but the direction of effect, for example whether the intervention effect is likely to be inflated, is unclear.
- Performance bias
  - We evaluated whether the knowledge of the allocated intervention could influence the outcome. In a placebo trial, patients and their healthcare providers do not know whether they receive the treatment or a placebo and so that knowledge cannot influence their behavior and the risk of performance bias is low. If people know that they are under observation, they may change their behavior (Hawthorne effect) and the risk of performance bias is high.
- Detection bias
  - The type of outcome assessment is critical in obesity prevention research. We evaluated whether the outcome assessor or the method of outcome assessment could be influenced by the participants and modified due to knowledge of the allocated intervention. Only studies using objective measures of physical activity were eligible for inclusion in the review. In the absence of suspected detection bias these studies were assessed as low risk of bias. Energy consumption may be measured through various self-report methods that differ in their reliability. Food frequency questionnaires (FFQ) are practical, well-suited to ranking individuals but estimated nutrient intakes derived from FFQs are imprecise. Moreover, FFQs are vulnerable to misreporting of food consumption that may in part be influenced by the knowledge of the intervention. Thus, these were considered high risk of bias. Food diaries were considered low risk of bias only in the presence of validations (e.g., checks) and the absence of suspected sources of detection bias. Twenty-four-hour dietary recall were considered low risk of bias. In studies where participants / outcome assessors were blind to the intervention allocation (placebo condition), detection bias was determined to be low risk.
- Attrition bias
  - We evaluated incomplete outcome data and, in particular, imbalances in follow-up data and selective dropout that is likely to be associated with the intervention. Attrition bias is suspected when there are systematic differences between treatment groups (pre vs post, intervention vs control) in withdrawals from the study. Studies with no missing data and loss to follow up and studies reporting intention to treat data were considered low risk of bias.

- Other sources of bias
  - We captured any additional aspects that could potentially affect the validity of the reported results in individual studies.

The sources of bias are compatible with the Cochrane Risk of Bias tool and are applicable to studies with historic or concurrent comparator. Reporting bias of individual studies was not assessed given that this review is interested in very specific outcomes. However, reporting bias were assessed across studies in the form of publication bias.

The results of the critical appraisal is documented in the figure below.
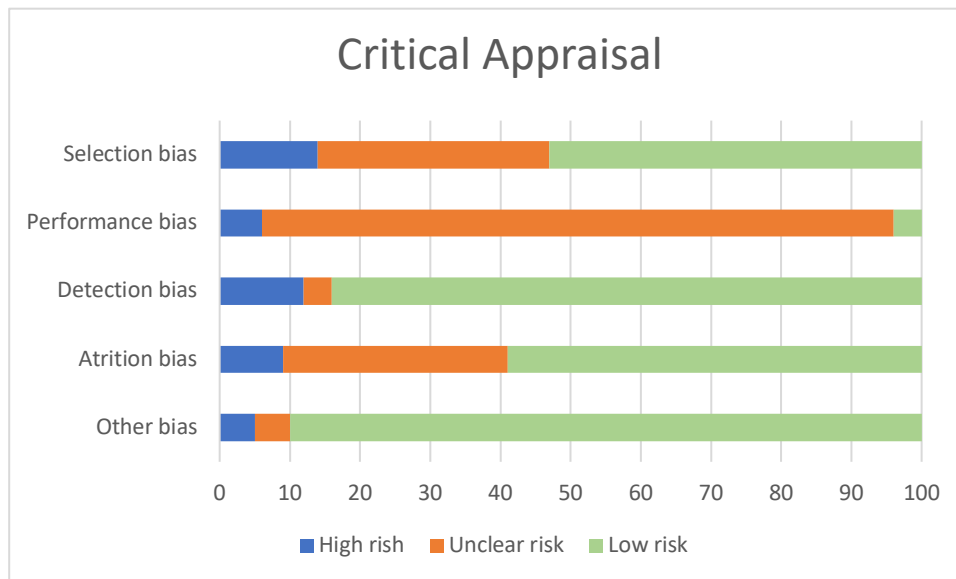
The results of the risk of bias assessments for the individual studies were incorporated into the quality of evidence summary (study limitation).