FOOD STANDARDS Australia New Zealand
Te Mana Kounga Kai – Ahitereiria me Aotearoa

19 November 2022

Supporting document 4 – Assessment of caffeine and sports performance

P1056 – Caffeine review

Executive summary

The objective of Proposal P1056 is to review permissions for caffeine in the Australia New Zealand Food Standards Code. The assessment described in this Supporting Document considers whether caffeine intake enhances sports performance.

FSANZ assessed evidence from human trials investigating the impact of caffeine intake on time trial performance in sports including cycling, running, rowing and swimming. Forty publications representing 39 studies and 42 pairwise comparisons were included. Eligible studies were mostly crossover trials with treatment order (caffeine or placebo) randomised, although most did not state a randomisation method.

Mean effect estimates from meta-analyses indicate that caffeine intake is associated with an observed faster time trial performance when compared to placebo. Using a standard scale of effect size, a small magnitude of effect was demonstrated by a meta-analysis of all 42 pairwise comparisons, using a pooled 674 participants, and a caffeine dose range of 1.25–9 mg per kilogram body weight (mg/kg BW).

We did not find a relationship between caffeine dose and effect size. Separate meta-analyses of time trial performance with the following caffeine dose ranges/doses gave similar effect sizes: 1.25–3 mg/kg BW; 4–6 mg/kg BW; 5 mg/kg BW; and 6 mg/kg BW. A meta-analysis was not conducted for caffeine doses >6 mg/kg BW as only two studies were available.

The level of certainty of the body of evidence is low. This means that our confidence in the effect estimate is limited; the true effect may be markedly different from the estimated effect. Our certainty in the evidence is reduced to low due to risk of bias and indirectness. Most studies did not state a randomisation method and some studies were not randomised. Most of the study participants were young adult males who were trained athletes with a high aerobic capacity. We have a low level of certainty that the effect size from analysis of the current data will apply to females, other age groups (children, adolescents, and older adults), untrained or unfit people, or for sports where performance is not correlated with aerobic exercise capacity.

We conclude with a low level of certainty that caffeine has a small beneficial effect, that is, a faster time trial performance after caffeine intake when compared to placebo. The lowest and highest dose level at which a small beneficial effect is observed lies within the range 1.25–3 mg/kg BW and at 6 mg/kg BW, respectively.

1. Assessment

1.1 Background

In 2004, the World Anti-Doping Agency (WADA) removed caffeine from the list of banned substances in sport. Caffeine is currently on the WADA Monitoring Program to detect potential patterns of misuse in sport. As for any substance, caffeine can be added to the WADA Prohibited List in future if it satisfies any two of the following three criteria: 1. It has the potential to enhance or enhances sport performance; 2. It represents an actual or potential health risk to the Athlete; or, 3. It violates the spirit of sport (as defined in the WADA Code).

In an umbrella review, Grgic et al. (2020) systematically reviewed published meta-analyses of the effects of caffeine on sports performance. Eleven reviews with a total of 21 meta-analyses were included, reflecting seven domains of sports performance: aerobic endurance; muscle strength; muscle endurance; anaerobic power; vertical jump height; exercise speed; and, short-term high-intensity exercise. Grgic et al. (2020) concluded that a performance-enhancing effect of caffeine is likely on muscle endurance, muscle strength, anaerobic power, and aerobic endurance in young male athletes. The level of certainty in the body of evidence in relation to all apparently healthy individuals of both sexes and all ages, varied depending on the meta-analysis. For aerobic endurance, the level of certainty of the meta-analytical evidence was categorised as low for four meta-analyses, very low for two, and moderate for three.

However, these findings have two main limitations. First, three of these four outcomes (muscle strength, anaerobic power, and aerobic endurance) included meta-analyses with 95% confidence intervals (CIs) that cross the line of no effect i.e., the uncertainty in the mean effect gives rise to a possible null or negative effect. The second main limitation is the substantial overlap or multiple-counting of primary research across individual meta-analyses of the same outcome. The resulting over-weighting of some studies precludes us from summarising the effect on an outcome by pooling the effect sizes from each meta-analysis, and has the potential to falsely indicate a more consistent effect (e.g., as indicated by the largely overlapping 95% CIs). If the overlap is due to selection or publication bias, the overlap also has the potential to skew the effect sizes in one direction.

Because of these limitations, we conducted an independent analysis using primary research that avoids multiple-counting of individual studies, to clarify the direction and magnitude of effect. This will also reduce any false inconsistency of meta-analytical findings presented by the Grgic et al. (2020) umbrella review that could be due to different methodological decisions of the systematic reviews, for example in deciding the research question, search strategy, and statistical analysis.

1.2 Objectives of the assessment

This assessment aims to answer:
Does caffeine intake have a beneficial effect on aerobic exercise performance? If so, at what dose range?

1.3 Methods

1.3.1 Question formulation

We conducted a scoping review and meta-analysis of the effect of caffeine intake on the duration to complete time trials, as a proxy for aerobic exercise performance. We selected

this outcome because it has been reported in published systematic reviews and meta-analyses that: (i) caffeine is more likely to have a larger beneficial effect on aerobic exercise performance compared to other domains of sports performance (Grgic et al. 2020), (ii) caffeine may be highly effective for enhancing time trial performance (Goldstein et al. 2010), and (iii) high users of caffeine among sports people are those competing in aerobic related sports (i.e., cycling, rowing, and athletics; Aguilar-Navarro et al. 2019).

1.3.2 Study selection and inclusion criteria

For efficiency, we replaced a literature search using databases with a hand search of the reference lists of six key publications, and imposed narrower criteria (Table 1). We identified primary research by screening individual publications that met the inclusion criteria of the five systematic reviews cited by Grgic et al. (2020) in relation to aerobic endurance (Doherty & Smith 2004, Conger et al. 2011, Ribeiro et al. 2017, Southward et al. 2018, and Shen et al. 2019). Consequently, the eligibility criteria were dictated by that of Grgic et al. (2020) with exception for the additional limitations we imposed for intervention, outcome, time, and study design (see Table 1), and being full-text peer-reviewed publications in the English language.

Table 1 PICOS criteria for study selection

Population	Apparently healthy individuals of both sexes and all ages.
Intervention	Any acute study[1] examining the effects of caffeine[2] ingestion on exercise performance.
Comparator	Placebo (provided that the effects of caffeine could be isolated).[3]
Outcome	Duration to complete time trial of a set task testing aerobic exercise performance.[4]
Study design	Placebo controlled trials.

[1] We adopted this from Grgic et al. (2020) and interpret this as caffeine intake before and/or during the time trial test.
[2] We excluded studies using food sources of caffeine (e.g. coffee, tea) as such studies preclude attribution of any observed effects to caffeine.
[3] We included studies that combined taurine with caffeine only if the placebo and intervention contained the same taurine concentration.
[4] We categorised time trials ≥5 minutes as aerobic exercise.

1.3.3 Data extraction, conversion, and analyses

Information on study characteristics was extracted from eligible studies, including study design, sample size and characteristics, placebo and supplementation protocol, and testing protocol. Outcome data were extracted from eligible studies and included means, standard deviations (SD) or standard errors (SE), and sample sizes for the intervention and comparator conditions. If outcome data were present only in graphs, the relevant data were extracted using the online program WebPlotDigitizer Version 4.4.

Where multiple pairwise comparisons were available, only the arms meeting our inclusion criteria were extracted. Additionally, we extracted the data for placebo versus caffeine in preference to carbohydrate versus caffeine with carbohydrate (Acker-Hewitt et al. 2012), placebo versus caffeine in preference to instant decaffeinated coffee versus instant coffee (Hodgson et al. 2013), placebo versus caffeine in preference to a substance (ephedrine, sodium bicarbonate, nitrate) versus caffeine with the same substance (ephedrine, sodium bicarbonate, nitrate) (Bell et al. 2002, Carr et al. 2011, and Glaister et al. 2015, respectively), and carbohydrate versus caffeine with carbohydrate in preference to fat versus caffeine with fat (Jacobson et al. 2001).

We extracted data for cycling in preference to handcycling (Graham-Paulson et al. 2016), placebo in preference to a 'no supplement control' condition (Gonçalves et al. 2017), a caffeine dose timing of 1 hour pre-exercise in preference to a timing to coincide peak serum caffeine concentrations (as determined from caffeine profiling) with onset of the time trial

(Skinner et al. 2013), and a caffeine dose timing of 1 hour pre-exercise in preference to during the time trial (Cox et al. 2002).

Results were averaged for studies making multiple observations of the same outcome, for example due to repeated time trials on the same sample (Astorino 2011, Astorino 2012a, and Astorino 2012b).

Standard deviations of group means, where not reported, were obtained from SEs of group means, using:

$$SD = SE \times \sqrt{N}$$

Two or three intervention arms of different caffeine doses were used in six studies (Cohen et al. 1996, Desbrow et al. 2009, Desbrow et al. 2012, Guest et al. 2018, Kovacs et al. 1998, and Skinner et al. 2010). Two intervention arms, representing a single or split dose of the same amount, were used in one study (Conway et al. 2003). To avoid unit-of-analysis error, we selected the intervention arm using the lowest caffeine dose or, for Conway et al. (2003), the single dose. Last, all outcome data (time) were standardised to minutes.

Meta-analyses were conducted in Stata version 16.1, developed by StataCorp LLC (TX, USA). Meta-analyses were performed using a random effects model using Hedge's *g* for effect size which adjusts for small sample bias. The I^2 statistic was used to assess heterogeneity of results between interventions. It describes the "percentage of total variation across studies that is not due to chance". I^2 values of 25%, 50% and 75% can be interpreted as indicating low, medium, and high heterogeneity, respectively (Higgins et al. 2003). Effect sizes using Hedge's *g* are described as: 0.1, very small; 0.2, small; 0.5, medium; 0.8, large; 1.2, very large; and, 2.0, huge (Sawilowsky 2009).

A meta-regression and subgroup analysis was planned *a priori* for dose, to investigate if there was a dose-response relationship and at what range (if any) caffeine intake has a beneficial effect on aerobic exercise performance. All other decisions were made *a posteriori*, including the following. Based on the interventions' caffeine doses range (from 1.25 to 9 mg/kg BW) we arbitrarily divided the studies into three subgroups of equal dose range (>0 to ≤3; >3 to ≤6; and >6 to ≤9 mg/kg BW). Only two studies fell in the dose range >6 to ≤9 mg/kg BW (Wemple et al. 1997 and Hunter et al. 2002), so we excluded these from the subgroup analysis. Based on results of the subgroup analysis, we conducted separate analyses on studies using a dose of 5 mg/kg BW and 6 mg/kg BW. Last, after completing study summaries, we assessed the impact of industry involvement as a potential moderator.

1.3.4 Strengths and limitations of our assessment

See Appendix 1, Table 6.

1.4 Results

1.4.1 Included studies and study characteristics

Of the publications meeting the inclusion criteria of existing systematic reviews, we retrieved 95 publications: six from the 40 publications included by Doherty & Smith (2004); 13 from 49 publications included by Conger et al. (2011); seven from 13 publications included by Ribeiro et al. (2017); 35 from 44 publications included by Southward et al. (2018); and, 34 from 40 publications included by Shen et al. (2019).

During full-text screening, we removed 40 duplicates and a further 15 publications as follows: four publications of time trials where duration (time) was not the dependent variable, as we

needed to pool results (Laurence et al. 2012, McNaughton et al. 2008a and 2008b, and Stadheim et al. 2014); one publication because the required data could not be directly extracted or calculated from the publication (Berglund & Hemmingsson 1982); four publications because the comparator, but not intervention, received dextrose powder or glucose (Walker et al. 2008, Anderson et al. 2000, Bridge & Jones 2006, and Bruce et al. 2000); two publications involving carbohydrate depletion the day before the time trial (Silva-Cavalcante et al. 2013) or negative energy balance for two days before the time trial (Slivka et al. 2008) reflecting additional interventions beyond caffeine, as well as being non-acute studies; one publication because the time trial duration was <5 minutes (Kilding et al. 2012); one publication because the caffeine source of the intervention was coffee (Church et al. 2015); and, two publications of a five- or six-day (i.e. not acute) study (Irwin et al. 2011 and Dean et al. 2009).

A final 40 publications, representing 39 studies and 42 pairwise comparisons, were included in our review and meta-analysis. Tables 2 to 4 present a summary of study characteristics. Table 5 presents outcome data.

Study design
All 39 studies used a crossover or similar study design, with the treatment order randomised in 33 studies. Only five studies state the method of randomisation (Guest et al. 2018, Potgieter et al. 2018, Quinlivan et al. 2015, Skinner et al. 2010, and Skinner et al. 2013). Five do not appear to be randomised (one being administered in a semi-counterbalanced fashion, another where treatment order was assigned, and three made no reference to randomisation) and one study is described as semi-randomised. Thirty-two and five studies were double- and single-blinded, respectively, one study appears to be single-blinded for the conditions relevant to the current assessment (Cox et al. 2002), and one made no mention of blinding (van Nieuwenhoven et al. 2005).

Studies reporting results of subgroups
Astorino et al. 2011 and Astorino et al. 2012a report on the same study. Three of the 39 studies report the results of two independent subgroups; two studies used trained and recreationally active subgroups (Astorino et al. 2011, Astorino et al. 2012a, and O'Rourke et al. 2008) and one study used AA homozygotes and C allele carriers (including both heterozygotes and CC homozygotes) subgroups (Womack et al. 2012). Figures 1 to 4 and 7 include data from all 42 pairwise comparisons from 39 studies.

Sample characteristics
The sample size of the 42 pairwise comparisons was: median n=10.5; mean n=16; and, range n=6–101. Of the pooled n=674 participants, 8% (n=57) were female. Studies (39) were conducted in Australia (13), USA (7), United Kingdom (6), Brazil (3), Canada (3), The Netherlands (2), Norway (2), South Africa (2), and Belgium (1). None were conducted in New Zealand. The mean age of participants was from 20 to 41 years. Publications often described training status using terms such as 'trained', 'endurance-trained', 'highly trained', 'recreational', and 'active'; sometimes with a training frequency or volume. Categorisation by investigators is inconsistent across studies. We categorised training status using VO_{2max} or VO_{2peak} which reflects the maximum rate of oxygen consumption. We used a sample mean of ≥55 mL/kg/min as the threshold to categorise a sample as 'highly aerobically trained'. This threshold was applied to male and females, and all ages. Our categorisation of studies based on sample means was: 'not trained' (5); 'trained' (21); one of two subgroups from one study 'not trained' and the other 'trained' (Astorino et al. 2011 and Astorino et al. 2012a); and, uncategorised (12 studies did not report VO_{2max} or VO_{2peak}). There is a high level of agreement between published description of training status and our categorisation based on VO_{2max} or VO_{2peak}; a small number differed (six of 27 studies; three were re-categorised as 'trained' and three were re-categorised as 'not trained'). We categorised the studies as follows: 55% of 39 studies are 'trained'; 31% are uncategorised; and, 14% are 'not trained'.

At the individual participant level (pooled n=674) this approximates to 36% (n=245) 'trained', 38% (n=253) uncategorised, and 26% (n=176) 'not trained'.

Caffeine intervention and placebo comparator
Caffeine was typically provided as a single dose only. Our analyses used data from the single dose condition in lieu of a split dose option for one study. That is, meta-analyses represented in Figures 1 to 4 and 7 include 33 studies (or 36 pairwise comparisons) that used a single caffeine dose, and six studies that split the caffeine dose across two or more time points. Of the pairwise comparisons using a single dose, caffeine was consumed on average ~80 minutes before the time trial. Caffeine dose ranged from 1.25 to 9 mg/kg BW. Analyses mean dose=4.8 mg/kg BW and median dose=5.0 mg/kg BW (except for Figures 5 and 6). Caffeine was provided in the form of a pill, capsule or tablet (27 studies), a beverage (11), or a gel (1). The form of placebo comparator matched that of the intervention condition.

Participants' pre-trial food intake
Seventeen studies' protocols provided both intervention and comparator conditions with a carbohydrate and/or electrolyte solution, carbohydrate-containing gel, or carbohydrate-containing concentrated beetroot juice 'shot'. Participants from 20 studies had consumed a meal, snack or liquid meal (self-selected or standardised, self-supplied or provided; these meals being different or additional to the solutions, gel, and shot discussed above) within 3 hours of the experiment. Publications of twelve studies do not mention the consumption of any food (via a solution, gel, shot, meal, snack, or liquid meal) in any condition (i.e. mutually exclusive with the 17 and 20 studies referred to above). Of exception to this is five studies that provided carbohydrate (presumably a small amount) to the comparator condition only via placebo capsules; nine studies' comparator condition included a placebo capsule containing maltodextrin, glucose, lactose, dextrose, or sucrose (of these, four studies included a carbohydrate-containing 'shot' or a meal in both the comparator and intervention condition). In most studies, participants were asked to avoid caffeine, alcohol and strenuous exercise for 24 or 48 hours prior to the time trial and to maintain the same diet for the day before each trial.

Time trials
Most studies used a cycle time trial (28 studies or ~70%), followed by run (4), row (3), double-pole cross-country ski (2), triathlon (1), or swim (1). The duration of the time trials ranged from 00:05:42 to 02:35:36 (hour:minute:second), with mean=00:40:40 and median=00:31:48 (time of the pooled intervention conditions from pairwise comparisons represented in Figures 1 to 4 and 7). The duration between trials in the same study ranged from 48 hours to 14 days. Familiarisation time trials, which aim to accustom participants to the procedures and minimise any potential learning or anxiety effects, were not completed in one third of the studies. Of these, only two studies investigated the effects of treatment order on performance (time); one found no treatment effect (Conway et al. 2003) and a second statistically controlled for potential learning effect (Guest et al. 2018).

Adverse effects
The safety of caffeine intake is addressed in detail in Supporting Document 1. Of the studies included in the nutrition assessment only, twelve studies reported adverse effects associated with caffeine, described as anxiety, nausea, mild tremor, feeling hyperactive or on edge, tachycardia, elevated or irregular heartbeat, impaired concentration, difficultly sleeping, mild gastrointestinal upset, light-headedness, headache, muscle cramping. One participant was unable to complete the time trial due to nausea and nervousness from a 6 mg/kg BW caffeine dose (Conway et al. 2003). Four studies reported: no adverse effects (2); no gastrointestinal complaints (1); and, no serious symptoms of gastrointestinal distress (1). The remaining 23 studies made no reference to adverse effects.

Industry involvement and trial registry

One quarter (11 studies) had industry involvement such as funding or in-kind contribution, or via employment. Twenty-six studies declared no conflicting interests or industry funding. Two studies provided no information regarding potential conflicts of interest. None were registered trials (e.g. via a clinical trial registry or similar).

1.4.2 Effect of caffeine on duration to complete time trials

Meta-analysis demonstrates that caffeine dose is associated with an observed faster time trial performance when compared to placebo (Figure 1). The magnitude of mean effect is small (-0.26), with the 95% CI ranging from a small to very small effect (-0.37, -0.16), and significant (P<0.01). There is high certainty (≥95%) in the absence of a neutral (i.e. no effect) or positive (i.e. faster in placebo) mean effect. Heterogeneity is negligible (I^2=0%) i.e., the variation of the mean effect size across the studies is very low and is not due to chance.

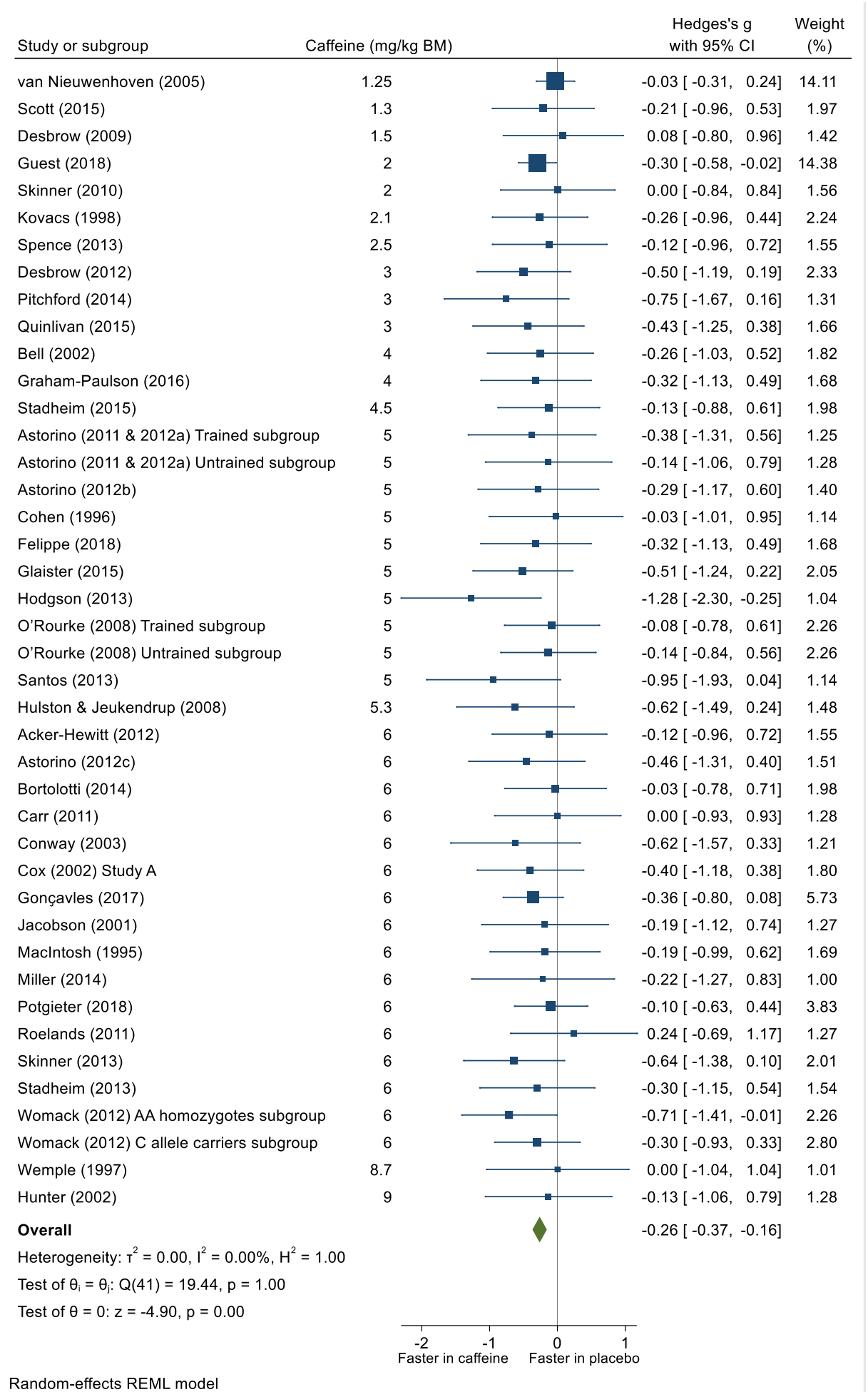
Study or subgroup
Caffeine (mg/kg BM)
Hedges's g with 95% CI
Weight (%)
van Nieuwenhoven (2005) 1.25 -0.03 [-0.31, 0.24] 14.11
Scott (2015) 1.3 -0.21 [-0.96, 0.53] 1.97
Desbrow (2009) 1.5 0.08 [-0.80, 0.96] 1.42
Guest (2018) 2 -0.30 [-0.58, -0.02] 14.38
Skinner (2010) 2 0.00 [-0.84, 0.84] 1.56
Kovacs (1998) 2.1 -0.26 [-0.96, 0.44] 2.24
Spence (2013) 2.5 -0.12 [-0.96, 0.72] 1.55
Desbrow (2012) 3 -0.50 [-1.19, 0.19] 2.33
Pitchford (2014) 3 -0.75 [-1.67, 0.16] 1.31
Quinlivan (2015) 3 -0.43 [-1.25, 0.38] 1.66
Bell (2002) 4 -0.26 [-1.03, 0.52] 1.82
Graham-Paulson (2016) 4 -0.32 [-1.13, 0.49] 1.68
Stadheim (2015) 4.5 -0.13 [-0.88, 0.61] 1.98
Astorino (2011 & 2012a) Trained subgroup 5 -0.38 [-1.31, 0.56] 1.25
Astorino (2011 & 2012a) Untrained subgroup 5 -0.14 [-1.06, 0.79] 1.28
Astorino (2012b) 5 -0.29 [-1.17, 0.60] 1.40
Cohen (1996) 5 -0.03 [-1.01, 0.95] 1.14
Felippe (2018) 5 -0.32 [-1.13, 0.49] 1.68
Glaister (2015) 5 -0.51 [-1.24, 0.22] 2.05
Hodgson (2013) 5 -1.28 [-2.30, -0.25] 1.04
O'Rourke (2008) Trained subgroup 5 -0.08 [-0.78, 0.61] 2.26
O'Rourke (2008) Untrained subgroup 5 -0.14 [-0.84, 0.56] 2.26
Santos (2013) 5 -0.95 [-1.93, 0.04] 1.14
Hulston & Jeukendrup (2008) 5.3 -0.62 [-1.49, 0.24] 1.48
Acker-Hewitt (2012) 6 -0.12 [-0.96, 0.72] 1.55
Astorino (2012c) 6 -0.46 [-1.31, 0.40] 1.51
Bortolotti (2014) 6 -0.03 [-0.78, 0.71] 1.98
Carr (2011) 6 0.00 [-0.93, 0.93] 1.28
Conway (2003) 6 -0.62 [-1.57, 0.33] 1.21
Cox (2002) Study A 6 -0.40 [-1.18, 0.38] 1.80
Gonçavles (2017) 6 -0.36 [-0.80, 0.08] 5.73
Jacobson (2001) 6 -0.19 [-1.12, 0.74] 1.27
MacIntosh (1995) 6 -0.19 [-0.99, 0.62] 1.69
Miller (2014) 6 -0.22 [-1.27, 0.83] 1.00
Potgieter (2018) 6 -0.10 [-0.63, 0.44] 3.83
Roelands (2011) 6 0.24 [-0.69, 1.17] 1.27
Skinner (2013) 6 -0.64 [-1.38, 0.10] 2.01
Stadheim (2013) 6 -0.30 [-1.15, 0.54] 1.54
Womack (2012) AA homozygotes subgroup 6 -0.71 [-1.41, -0.01] 2.26
Womack (2012) C allele carriers subgroup 6 -0.30 [-0.93, 0.33] 2.80
Wemple (1997) 8.7 0.00 [-1.04, 1.04] 1.01
Hunter (2002) 9 -0.13 [-1.06, 0.79] 1.28
Overall -0.26 [-0.37, -0.16]
Heterogeneity: τ^2 = 0.00, I^2 = 0.00%, H^2 = 1.00
Test of $\theta_i = \theta_j$: Q(41) = 19.44, p = 1.00
Test of θ = 0: z = -4.90, p = 0.00
-2
-1
0
1
Faster in caffeine
Faster in placebo
Random-effects REML model

Figure 1: Caffeine has a beneficial effect on time trial performance.
Description of trained versus untrained subgroups represents the investigators' categorisation. Our categorisation is the same for Astorino (2011 and 2012a) but different for O'Rourke (2008) which we classed as 'uncategorised' due to no VO_{2max} and VO_{2peak} being reported.

1.4.3 Publication bias

Publication bias is more common when there is industry involvement in most of the published studies. Funnel-plot symmetry is observed via visual inspection and statistical analyses i.e., no evidence of publication bias. Visual inspection of the funnel plot demonstrates symmetry around the reference line and smaller studies (not distinguished in Figures 2 and 3) are present in the non-significant (darkest) regions. A small amount of asymmetry with respect to the reference line is observed, towards the bottom third of the darkest region (with *p*-values larger than 10%). There is a chance we are missing some of the smaller trials with non-significant results, which would be consistent with the presence of publication bias. To assess this more formally, we used the Egger regression-based test for small-study effects. The *p*-value was non-significant (*P*=0.30) indicating no relationship between effect size and precision, i.e. funnel-plot symmetry.

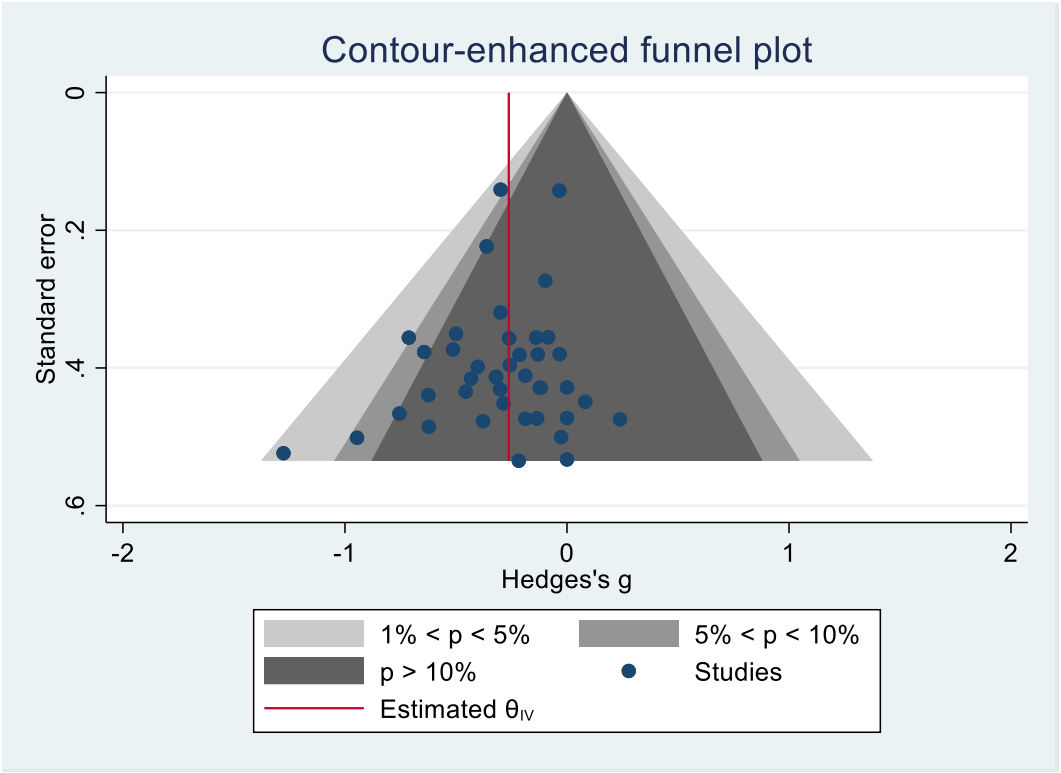

Figure 2: Counter-enhanced funnel plot does not demonstrate publication bias.

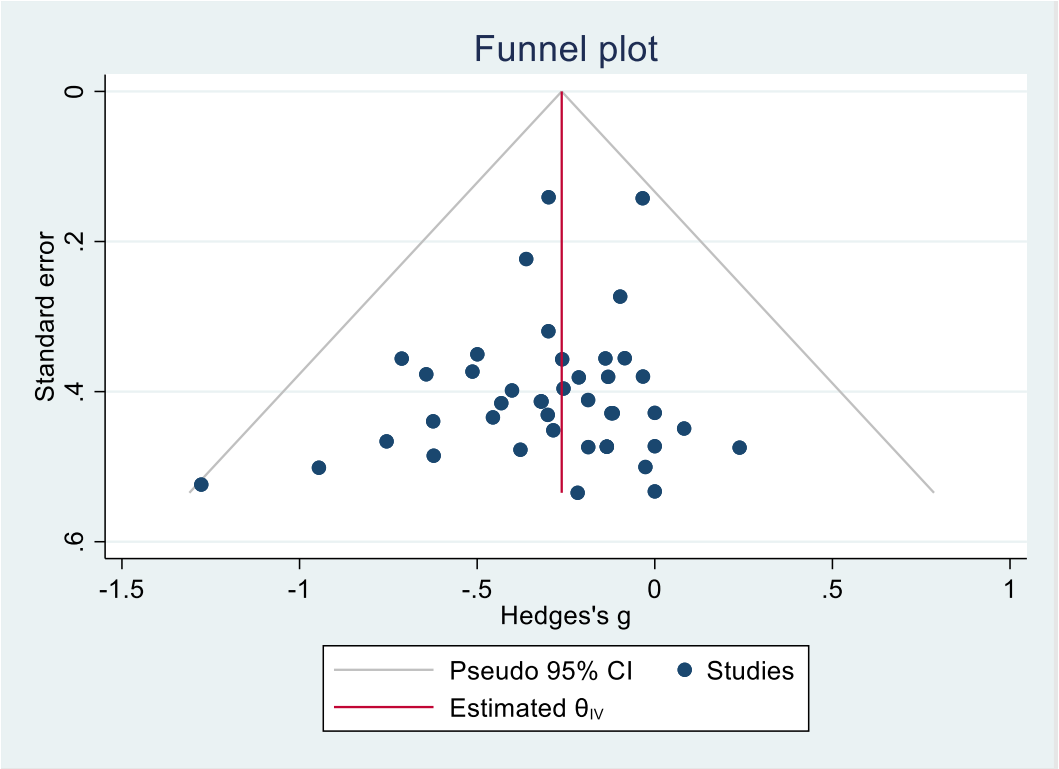

Figure 3: Funnel plot mostly illustrates symmetry around the reference line.

1.4.4 Moderator analyses: caffeine dose

Random-effects meta-regression demonstrates that caffeine dose is not associated with time trial performance (slope = -0.026, *P*=0.32). Subgroup analysis demonstrates that caffeine dose according to two categories, ≤3 mg/kg BW and >3 to ≤6 mg/kg BW, is not associated with the observed intervention effects on time trial performance. The magnitudes of effect of both groups did not differ significantly (*P*=0.32; Figure 4).

The lowest level at which a beneficial effect is observed in this data lies within the range 1.25–3 mg/kg BW. Subgroup analysis demonstrates that intake of a caffeine dose from 1.25 to 3 mg/kg BW is associated with an observed faster time trial performance when compared to placebo. The magnitude of mean effect is small (-0.21, 95% CI: -0.37, -0.04). There is high certainty (≥95%) in the absence of a neutral or negative mean effect (95% CI: -0.37, -0.04). Heterogeneity is not apparent (I^2=0%). Further analyses using a dose within 1.25–3 mg/kg BW was not possible due to an inadequate number of studies.

The highest level at which a beneficial effect is observed in this data occurs at 6 mg/kg BW. An analysis using the highest dose tested by 10 or more pairwise comparisons (6 mg/kg BW from 16 or 18 trials; see Figures 6a and 6b, respectively) demonstrates a faster time trial performance after caffeine intake when compared to placebo. The magnitude of mean effect is small when using 16 trials (-0.29, 95% CI: -0.47, -0.11; Figure 6a) or 18 trials (-0.28, 95% CI: -0.46, -0.11; Figure 6b). There is high certainty (≥95%) in the absence of a neutral or negative mean effect (95% CIs do not cross the line of no effect). Heterogeneity is negligible (both I^2=0%). Data from two extra studies are present in Figure 6b (Skinner et al. 2010 and Desbrow et al. 2012). These are excluded from other analyses (Figures 1–4 and 7); the 6 mg/kg BW dose data were excluded because we included only data from the lowest caffeine dose arms (2 mg/kg BW for Skinner et al. 2010 and 3 mg/kg BW for Desbrow et al. 2012) in studies with multiple intervention arms. A beneficial effect is also observed in pairwise comparisons using a 5 mg/kg BW dose (Figure 5).

We are unaware of studies (individual or pooled trials) demonstrating a dose-response relationship or caffeine doses at which a sports performance benefit starts and stops. An umbrella review by Grgic et al. (2020) could not establish an optimal dose and on this basis they recommend dose-response studies are needed. Souza et al. (2017) did not detect a significant association between caffeine dose from consumption of caffeine-containing energy drinks, and physical performance (P=0.21).

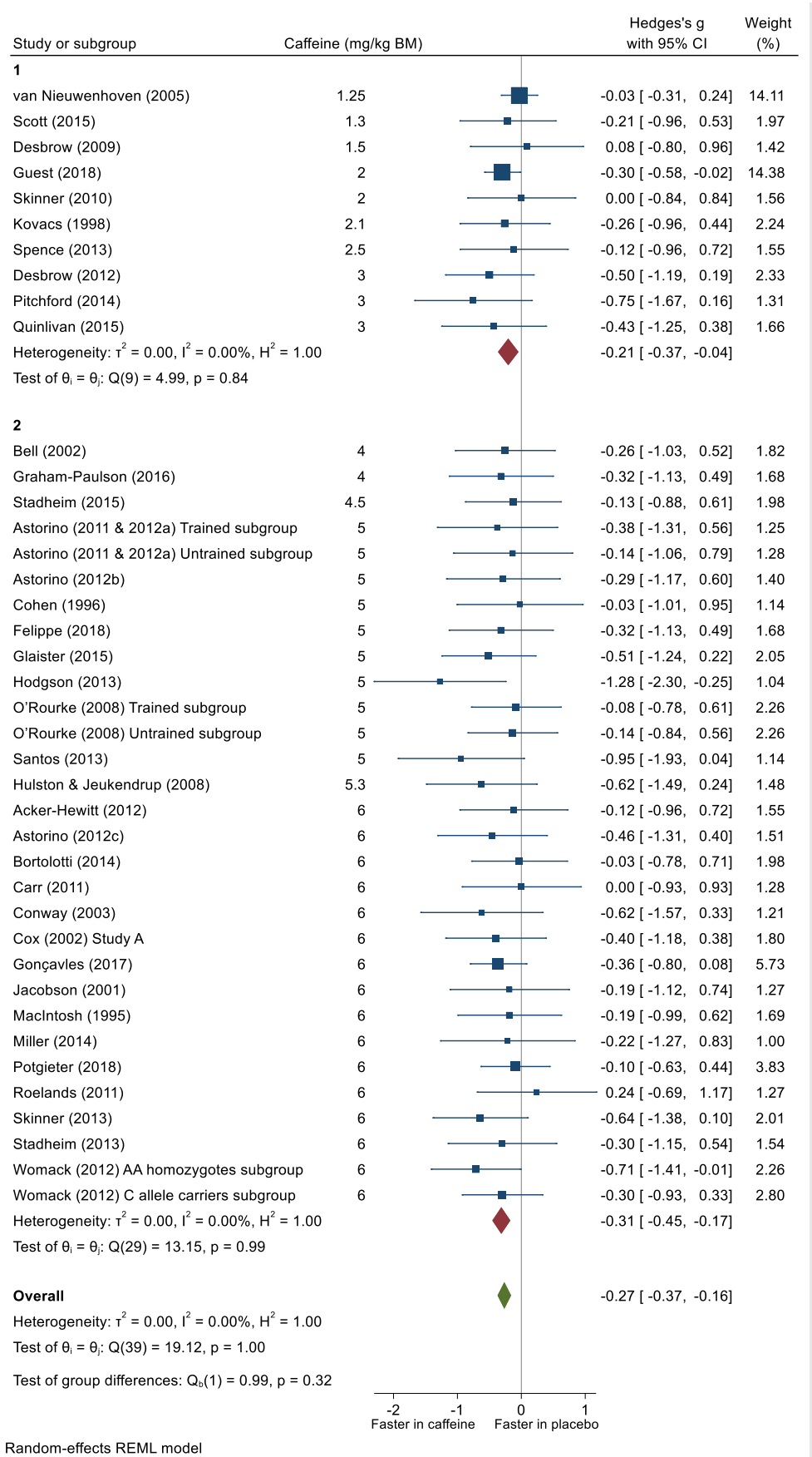

Figure 4: Two dose categories do not modify the effect of caffeine on time trial performance.

1, ≤3 mg/kg BW; 2, >3 to ≤6 mg/kg BW.

Description of trained versus untrained subgroups represents the investigators' categorisation. Our categorisation is the same for Astorino (2011 and 2012a) but different for O'Rourke (2008) which we classed as 'uncategorised' due to no VO_{2max} and VO_{2peak} being reported.

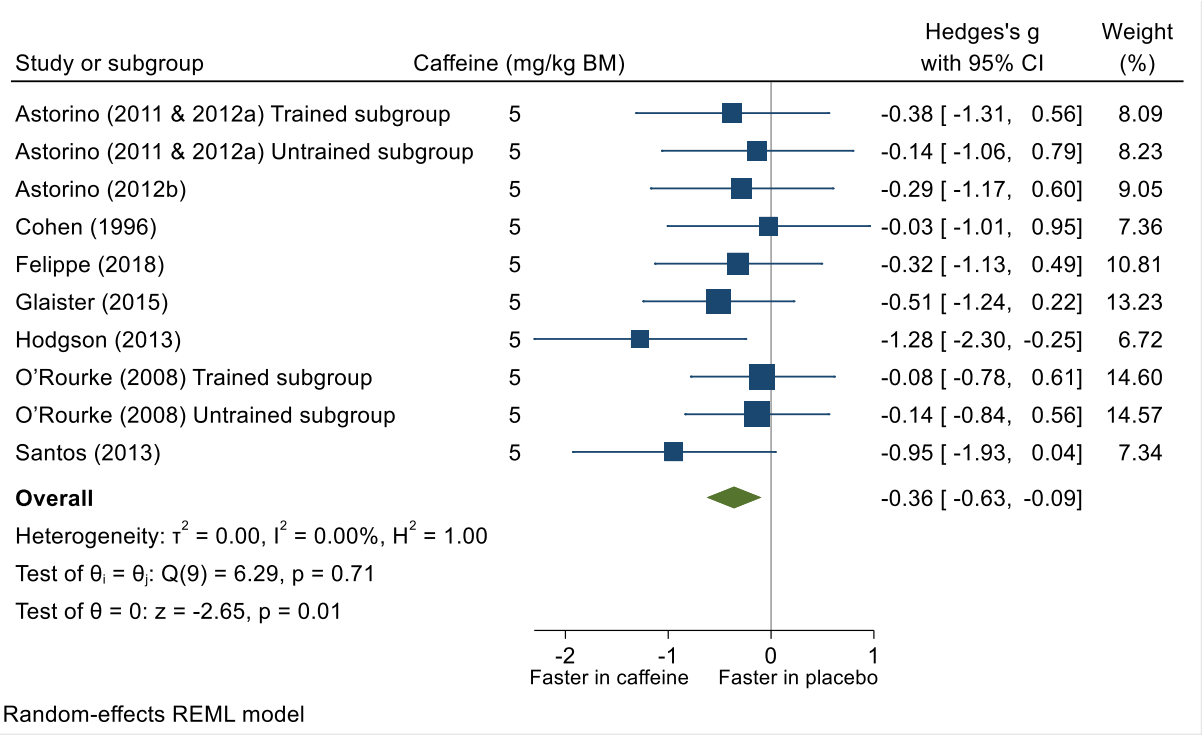

Figure 5: A 5 mg/kg body weight caffeine dose has a beneficial effect on time trial performance.

Description of trained versus untrained subgroups represents the investigators' categorisation. Our categorisation is the same for Astorino (2011 and 2012a) but different for O'Rourke (2008) which we classed as 'uncategorised' due to no VO_{2max} and VO_{2peak} being reported.

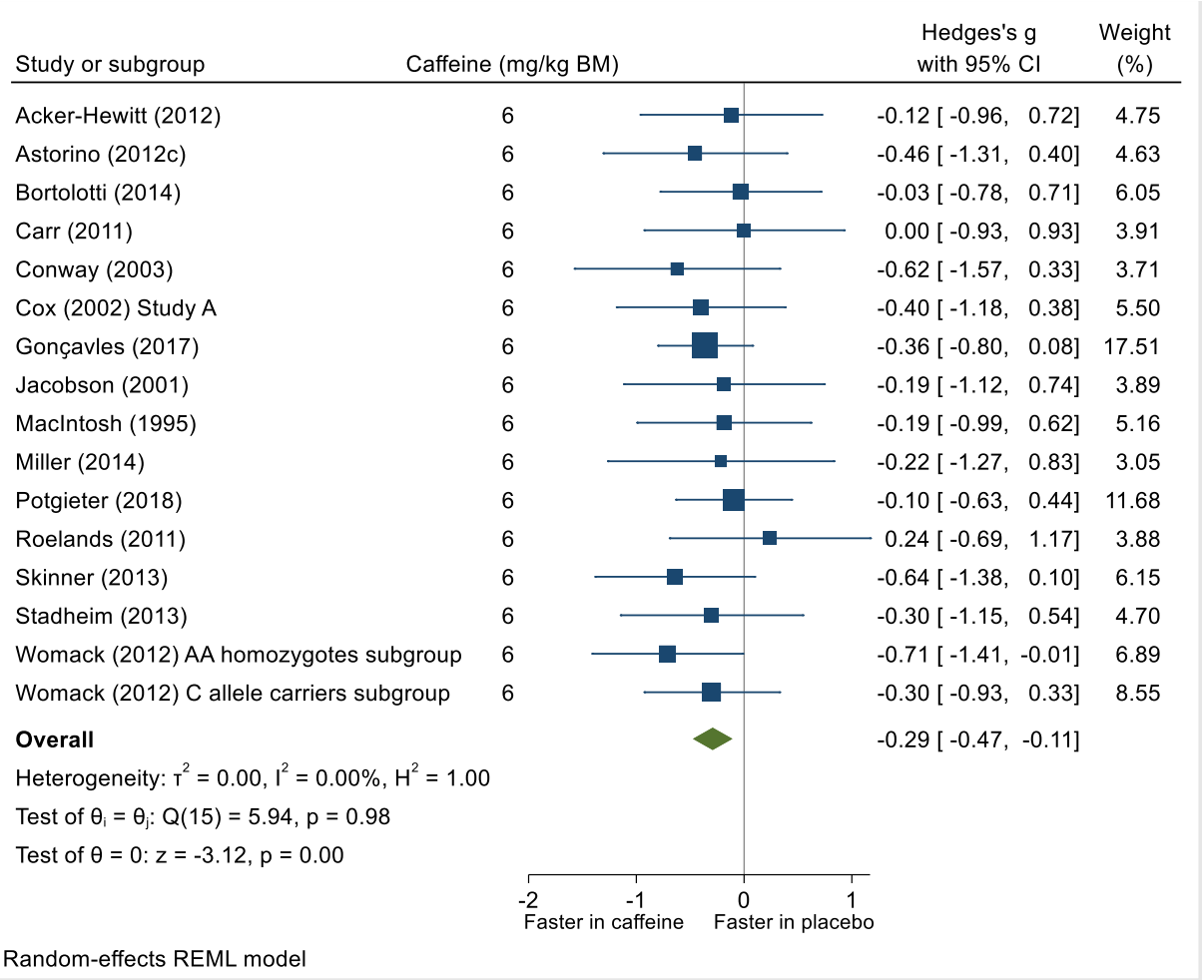

Figure 6a: A 6 mg/kg body weight caffeine dose has a beneficial effect on time trial performance (16 pairwise comparisons).

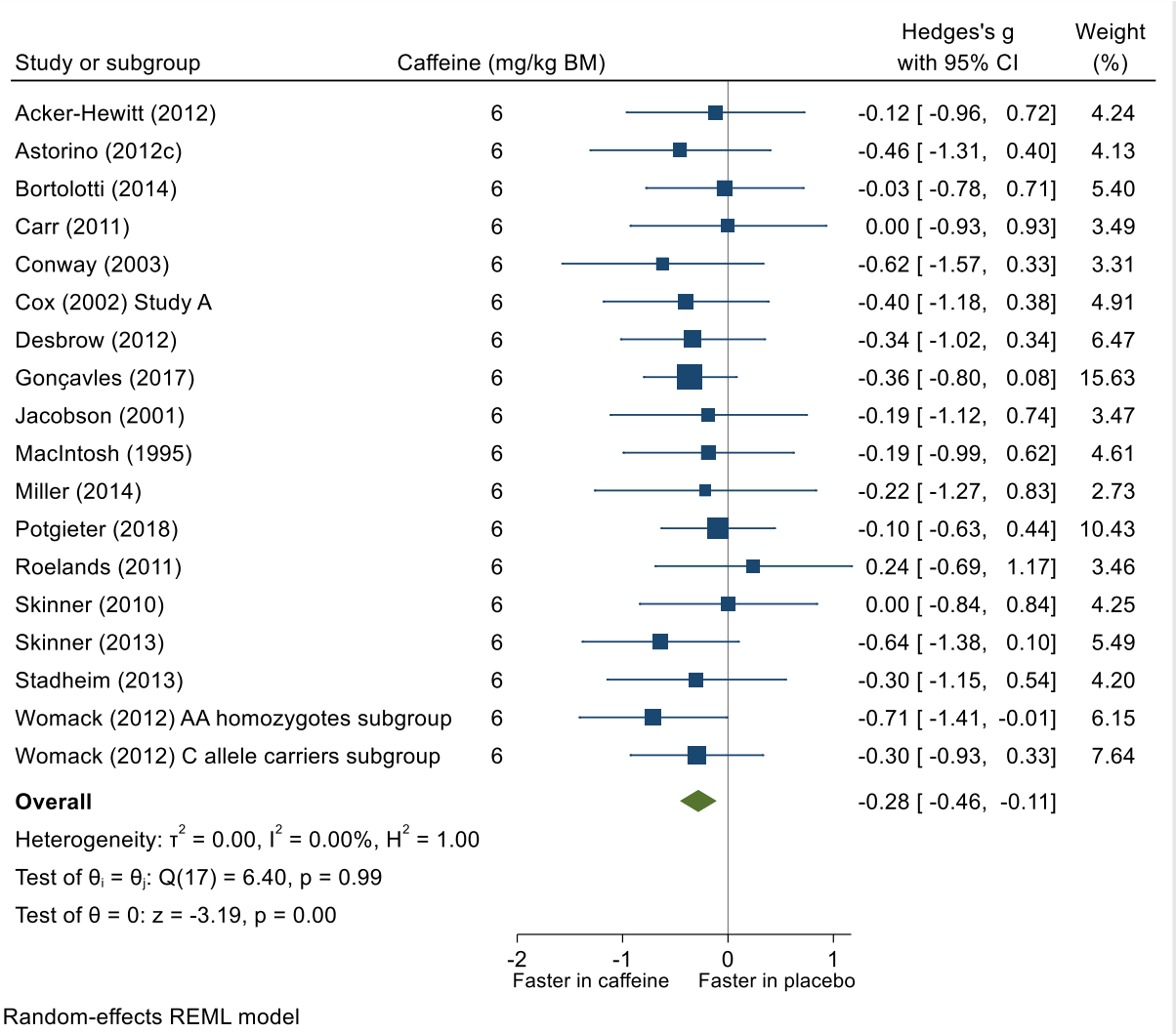

Figure 6b: A 6 mg/kg body weight caffeine dose has a beneficial effect on time trial performance (18 pairwise comparisons).

1.4.5 Moderator analyses: industry involvement

One quarter of studies had industry involvement, whether through declared interests or funding, which can be a source of bias. A subgroup analysis demonstrates that intervention effect is not modified by the studies' industry involvement. The magnitude of effect was slightly larger in the group of studies receiving industry links (summary effect size = -0.32 versus -0.24) but they did not differ significantly (*P*=0.47; Figure 7).

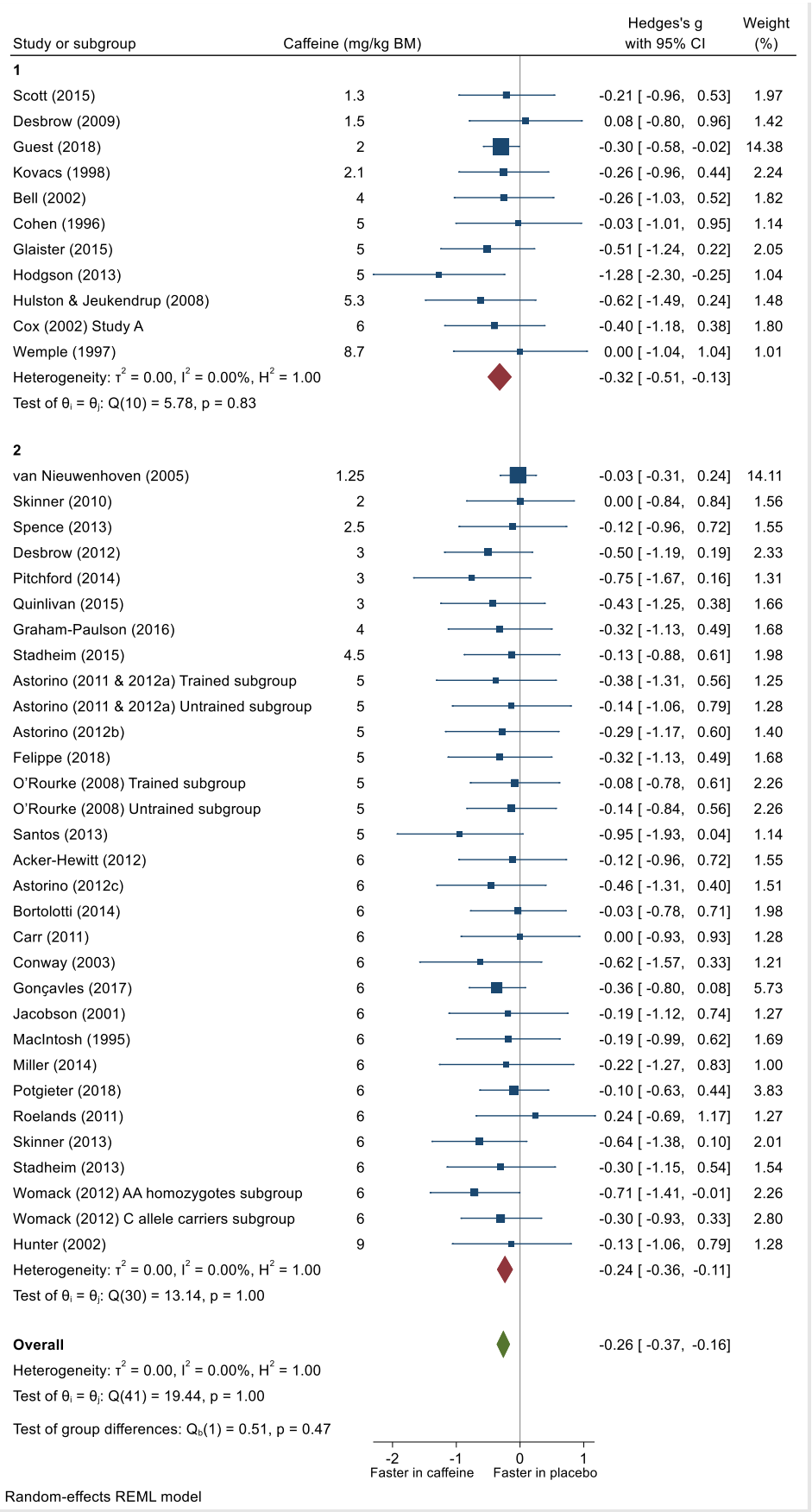

Figure 7: Industry involvement does not modify the effect of caffeine on time trial performance.

1, studies with industry involvement; 2, studies stating no industry involvement and those not declaring funding or

conflicts of interest.
Description of trained versus untrained subgroups represents the investigators' categorisation. Our categorisation is the same for Astorino (2011 and 2012a) but different for O'Rourke (2008) which we classed as 'uncategorised' due to no VO_{2max} and VO_{2peak} being reported.

1.4.6 Certainty in the evidence

Our level of certainty in the evidence is affected as follows: not downgraded due to minimal inconsistency, imprecision, and publication bias; not upgraded due to the absence of a large magnitude of effect, and a dose-response gradient; and, downgraded due to a risk of bias and indirectness. Our level of certainty in the evidence has not considered whether all residual confounding would decrease the magnitude of effect (in situations with an effect), which we did not assess.

Overall, our certainty in the evidence was downgraded by two levels to a low certainty, due to a risk of bias and indirectness. This means that our confidence in the effect estimate is limited: the true effect may be markedly different from the estimated effect. A low level of certainty is consistent with that of published meta-analyses. A summary of our results and discussion of the wider literature, with regards to certainty, is reported in Appendix 1.

1.4.7 Conclusions

We conclude with a low level of certainty that caffeine, compared to placebo, has a small beneficial effect on time trial performance. The lowest and highest level at which a small beneficial effect is observed lies within the dose range 1.25–3 mg/kg BW and at 6 mg/kg BW, respectively.

References

Acker-Hewitt TL, Shafer BM, Saunders MJ, Goh Q, Luden ND (2012) Independent and combined effects of carbohydrate and caffeine ingestion on aerobic cycling performance in the fed state. Appl Physiology Nutrition Metabolism 37:276–283

Anderson ME, Bruce CR, Fraser SF, Stepto NK, Klein R, Hopkins WG, Hawley JA (2000) Improved 2000-meter rowing performance in competitive oarswomen after caffeine ingestion. Int J Sport Nutr Exe 10:464–475

Astorino TA, Cottrell T, Lozano AT, Aburto-Pratt K, Duhon J (2011) Ergogenic effects of caffeine on simulated time-trial performance are independent of fitness level. J Caffeine Res 1:179–185

Astorino TA, Cottrell T, Lozano AT, Aburto-Pratt K, Duhon J (2012a) Effect of caffeine on RPE and perceptions of pain, arousal, and pleasure/displeasure during a cycling time trial in endurance trained and active men. Physiol Behav 106:211–217

Astorino TA, Cottrell T, Lozano AT, Aburto-Pratt K, Duhon J (2012b) Increases in cycling performance in response to caffeine ingestion are repeatable. Nutr Res 32:78–84

Astorino TA, Roupoli LR, Valdivieso BR (2012c) Caffeine does not alter RPE or pain perception during intense exercise in active women. Appetite 59:585–590

Bell DG, McLellan TM, Sabiston CM (2002) Effect of ingesting caffeine and ephedrine on 10-km run performance. Med Sci Sport Exer 34:344–349

Berglund B, Hemmingsson P (1982) Effects of caffeine ingestion on exercise performance at low and high altitudes in cross-country skiers. Int J Sports Med 3:234–236

Bortolotti, H, Altimari LR, Vitor-Costa M, Cyrino ES (2014) Performance during a 20-km cycling time-trial after caffeine ingestion. J Int Soc Sport Nutr 11:45

Bridge CA, Jones MA (2006) The effect of caffeine ingestion on 8 km run performance in a field setting. J Sport Sci 24:433–439

Bruce CR, Anderson ME, Fraser SF, Stepto NK, Klein R, Hopkins WG, Hawley JA (2000) Enhancement of 2000-m rowing performance after caffeine ingestion. Medicine Sci Sports Exerc 32:1958–1963

Carr AJ, Gore CJ, Dawson B (2011) Induced alkalosis and caffeine supplementation: effects on 2,000-m rowing performance. Int J Sport Nutr Exe 21:357–364

Church DD, Hoffman JR, LaMonica MB, Riffe JJ, Hoffman MW, Baker KM, Varanoske AN, Wells AJ, Fukuda DH, Stout JR (2015) The effect of an acute ingestion of Turkish coffee on reaction time and time trial performance. J Int Soc Sport Nutr 12:37

Cohen BS, Nelson AG, Prevost MC, Thompson GD, Marx BD, Morris GS (1996) Effects of caffeine ingestion on endurance racing in heat and humidity. Eur J Appl Physiol O 73:358–363

Conger SA, Warren GL, Hardy MA, Millard-Stafford ML (2011) Does caffeine added to carbohydrate provide additional ergogenic benefit for endurance? Int J Sport Nutr Exe 21:71–84

Conway KJ, Orr R, Stannard SR (2003) Effect of a divided caffeine dose on endurance cycling performance, postexercise urinary caffeine concentration, and plasma paraxanthine. J Appl Physiol 94:1557–1562

Cox GR, Desbrow B, Montgomery PG, Anderson ME, Bruce CR, Macrides TA, Martin DT, Moquin A, Roberts A, Hawley JA, Burke LM (2002) Effect of different protocols of caffeine intake on metabolism and endurance performance. J Appl Physiol 93:990–999

Dean S, Braakhuis A, Paton C (2009) The effects of EGCG on fat oxidation and endurance performance in male cyclists. Int J Sport Nutr Exe 19:624–644

Desbrow B, Barrett CM, Minahan CL, Grant GD, Leveritt MD (2009) Caffeine, cycling performance, and exogenous CHO oxidation: a dose-response study. Medicine Sci Sports Exerc 41:1744–1751

Desbrow B, Biddulph C, Devlin B, Grant GD, Anoopkumar-Dukie S, Leveritt MD (2012) The effects of different doses of caffeine on endurance cycling time trial performance. J Sport Sci 30:115–120

Doherty M, Smith PM (2004) Effects of caffeine ingestion on exercise testing: a meta-analysis. Int J Sport Nutr Exe 14:626–646

Felippe LC, Ferreira GA, Learsi SK, Boari D, Bertuzzi R, Lima-Silva AE (2018) Caffeine increases both total work performed above critical power and peripheral fatigue during a 4-km cycling time trial. J Appl Physiol 124:1491–1501

Ferreira RES, Pacheco RL, de Oliveira Cruz Latorraca C, Riera R, Eid RG, Martimbianco ALC (2021) Effects of caffeine supplementation on physical performance of soccer players: systematic review and meta-analysis. Sports Health. 13:347-358

Glaister M, Pattison JR, Muniz-Pumares D, Patterson SD, Foley P (2015) Effects of dietary nitrate, caffeine, and their combination on 20-km cycling time trial performance. J Strength Cond Res 29:165–174

Goldstein ER, Ziegenfuss T, Kalman D, Kreider R, Campbell B, Wilborn C, Taylor L, Willoughby D, Stout J, Graves BS, Wildman R, Ivy JL, Spano M, Smith AE, Antonio J (2010) International society of sports nutrition position stand: caffeine and performance. J Int Soc Sports Nutr 7(1):5

Gonçalves L de S, Painelli V de S, Yamaguchi G, de Oliveira LF, Saunders B, da Silva RP, Maciel E, Artioli GG, Roschel H, Gualano B (2017) Dispelling the myth that habitual caffeine consumption influences the performance response to acute caffeine supplementation. J Appl Physiol 123:213–220

GRADE Working Group (2013). GRADE Handbook. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Schünemann H, Brożżżerk J, Guyatt G, Oxman A.

https://gdt.gradepro.org/app/handbook/handbook.htmlGraham-Paulson T, Perret C, Goosey-Tolfrey V (2016) Improvements in cycling but not handcycling 10 km time trial performance in habitual caffeine users. Nutrients 8:393

Grgic J, Grgic I, Pickering C, Schoenfeld BJ, Bishop DJ, Pedisic Z (2020) Wake up and smell the coffee: caffeine supplementation and exercise performance—an umbrella review of 21 published meta-analyses. Brit J Sport Med 54:681-688

Guest N, Corey P, Vescovi J, El-Sohemy A (2018) Caffeine, CYP1A2 genotype, and endurance performance in athletes. Medicine Sci Sports Exerc 50:1570–1578

Higgens JP, Thompson SG, Deeks JJ, Altman DG (2003) Measuring inconsistency in meta-analyses. BMJ 327(7414):557-560

Hodgson AB, Randell RK, Jeukendrup AE (2013) The metabolic and performance effects of caffeine compared to coffee during endurance exercise. PLOS One 8:e59561

Hulston CJ, Jeukendrup AE (2008) Substrate metabolism and exercise performance with caffeine and carbohydrate intake. Medicine Sci Sports Exerc 40:2096–2104

Hunter AM, Gibson A St C, Collins M, Lambert M, Noakes TD (2002) Caffeine ingestion does not alter performance during a 100-km cycling time-trial performance. Int J Sport Nutr Exe 12:438–452

Irwin C, Desbrow B, Ellis A, O'Keeffe B, Grant G, Leveritt M (2011) Caffeine withdrawal and high-intensity endurance cycling performance. J Sport Sci 29:509–515

Jacobson TL, Febbraio MA, Arkinstall MJ, Hawley JA (2001) Effect of caffeine co-ingested with carbohydrate or fat on metabolism and performance in endurance-trained men. Exp Physiol 86:137–144

Kilding AE, Overton C, Gleave J (2012) Effects of caffeine, sodium bicarbonate, and their combined ingestion on high-intensity cycling performance. Int J Sport Nutr Exe 22:175–183

Kovacs EMR, Stegen JHCH, Brouns F (1998) Effect of caffeinated drinks on substrate metabolism, caffeine excretion, and performance. J Appl Physiol 85:709–715

Laurence G, Wallman K, Guelfi K (2012) Effects of caffeine on time trial performance in sedentary men. J Sport Sci 30:1235–1240

MacIntosh BR, Wright BM (1995) Caffeine ingestion and performance of a 1,500-metre swim. Can J Appl Physiology 20:168–177

Maughan RJ, Burke LM, Dvorak J, et al. (2018) IOC consensus statement: dietary supplements and the high-performance athlete. Br J Sports Med 52**:**439-455

Miller B, O'Connor H, Orr R, Ruell P, Cheng HL, Chow CM (2014) Combined caffeine and carbohydrate ingestion: effects on nocturnal sleep and exercise performance in athletes. Eur J Appl Physiol 114:2529–2537

O'Rourke MP, O'Brien BJ, Knez WL, Paton CD (2008) Caffeine has a small effect on 5-km running performance of well-trained and recreational runners. J Sci Med Sport 11:231–233

Pitchford NW, Fell JW, Leveritt MD, Desbrow B, Shing CM (2014) Effect of caffeine on cycling time-trial performance in the heat. J Sci Med Sport 17:445–449

Potgieter S, Wright HH, Smith C (2018) Caffeine improves triathlon performance: a field study in males and females. Int J Sport Nutr Exe 28:228–237

Quinlivan A, Irwin C, Grant GD, Anoopkumar-Dukie S, Skinner T, Leveritt M, Desbrow B (2015) The effects of Red Bull energy drink compared with caffeine on cycling time-trial performance. Int J Sport Physiol 10:897–901

Ribeiro BG, Morales AP, Sampaio-Jorge F, Tinoco F de S, de Matos AA, Leite TC (2017) Acute effects of caffeine intake on athletic performance: a systematic review and meta-analysis. Revista Chilena De Nutrición 44:283–291

Roelands B, Buyse L, Pauwels F, Delbeke F, Deventer K, Meeusen R (2011) No effect of caffeine on exercise performance in high ambient temperature. Eur J Appl Physiol 111:3089–3095

Santos R de A, Kiss MAPDM, Silva-Cavalcante MD, Correia-Oliveira CR, Bertuzzi R, Bishop DJ, Lima-Silva AE (2013) Caffeine alters anaerobic distribution and pacing during a 4000-m cycling time trial. PLOS One 8:e75399

Sawilowsky SS (2009) New effect size rules of thumb. J Mod Appl Statistical Methods 8:597–599

Scott AT, O'Leary T, Walker S, Owen R (2015) Improvement of 2000-m rowing performance with caffeinated carbohydrate-gel ingestion. Int J Sport Physiol 10:464–468

Shen JG, Brooks MB, Cincotta J, Manjourides JD (2019) Establishing a relationship between the effect of caffeine and duration of endurance athletic time trial events: a systematic review and meta-analysis. J Sci Med Sport 22:232–238

Silva-Cavalcante MD, Correia-Oliveira CR, Santos RA, Lopes-Silva JP, Lima HM, Bertuzzi R, Duarte M, Bishop DJ, Lima-Silva AE (2013) Caffeine increases anaerobic work and restores cycling performance following a protocol designed to lower endogenous carbohydrate availability. PLOS One 8:e72025

Skinner TL, Jenkins DG, Coombes JS, Taaffe DR, Leveritt MD (2010) Dose response of caffeine on 2000-m rowing performance. Medicine Sci Sports Exerc 42:571–576

Skinner TL, Jenkins DG, Taaffe DR, Leveritt MD, Coombes JS (2013) Coinciding exercise with peak serum caffeine does not improve cycling performance. J Sci Med Sport 16:54–59

Slivka D, Hailes W, Cuddy J, Ruby B (2008) Caffeine and carbohydrate supplementation during exercise when in negative energy balance: effects on performance, metabolism, and salivary cortisol. Appl Physiology Nutrition Metabolism 33:1079–1085

Southward K, Rutherfurd-Markwick KJ, Ali A (2018) Correction to: The effect of ccute caffeine ingestion on endurance performance: a systematic review and meta-analysis. Sports Med 48:2425–2441

Souza DB, Coso JD, Casonatto J, Polito MD (2017) Acute effects of caffeine-containing energy drinks on physical performance: a systematic review and meta-analysis. Eur J Nutr 56:13–27

Spence AL, Sim M, Landers G, Peeling P (2013) A comparison of caffeine versus pseudoephedrine on cycling time-trial performance. Int J Sport Nutr Exe 23:507–512

Stadheim HK, Kvamme B, Olsen R, Drevon CA, Ivy JL, Jensen J (2013) Caffeine increases performance in cross-country double-poling time trial exercise. Medicine Sci Sports Exerc 45:2175–2183

Stadheim HK, Nossum EM, Olsen R, Spencer M, Jensen J (2015) Caffeine improves performance in double poling during acute exposure to 2,000-m altitude. J Appl Physiol 119:1501–1509

Stadheim HK, Spencer M, Olsen R, Jensen J (2014) Caffeine and performance over consecutive days of simulated competition. Medicine Sci Sports Exerc 46:1787–1796

van Nieuwenhoven MA, Brouns F, Kovacs EMR (2005) The effect of two sports drinks and water on GI complaints and performance during an 18-km run. Int J Sports Med 26:281–285

Walker GJ, Dziubak A, Houghton L, Prendergast C, Lim L, Bishop NC (2008) The effect of caffeine ingestion on human neutrophil oxidative burst responses following time-trial cycling. J Sport Sci 26:611–619

Wemple R, Lamb D, McKeever K (1997) Caffeine vs caffeine-free sports drinks: effects on urine production at rest and during prolonged exercise. Int J Sports Med 18:40–46

Womack CJ, Saunders MJ, Bechtel MK, Bolton DJ, Martin M, Luden ND, Dunham W, Hancock M (2012) The influence of a CYP1A2 polymorphism on the ergogenic effects of caffeine. J Int Soc Spot Nutr 9:7

Appendix 1

Certainty in the evidence

We used the internationally recognised approach, GRADE (Grading of Recommendations, Assessment, Development and Evaluation), to rate the quality of evidence and the level of certainty in the evidence and thus, of our conclusions.

Our level of certainty in the evidence is affected as follows: not downgraded due to minimal inconsistency, imprecision, and publication bias; not upgraded due to the absence of a large magnitude of effect, and a dose-response gradient; and, downgraded due to a risk of bias and indirectness. Our level of certainty in the evidence has not considered whether all residual confounding would decrease the magnitude of effect (in situations with an effect), which we did not assess.

Inconsistency
Heterogeneity was not apparent (I^2=0%) and there is a consistent overlap of confidence intervals (all overlap with each other, and only three from 42 pairwise comparisons do not cross the line of no effect; Hodgson et al. 2013, Guest et al. 2018, and the AA homozygotes subgroup of Womack et al. 2012). Point estimates were ≥0.0 in about one in five comparisons representing no effect or a small disbenefit (range: -0.03 to 0.24 in eight from 42 comparisons), between 0.0 and -0.5 in about three from five comparisons representing a small to moderate magnitude of effect or greater in favour of the caffeine intervention (range: -0.08 to -0.43 in 24 from 42 comparisons), and ≤-0.5 in about one in five comparisons representing a moderate magnitude of effect or greater in favour of the caffeine intervention (range: -0.46 to -1.28 in ten from 42 comparisons). The heterogeneity by I^2, confidence intervals and point estimates indicate some variation as well as similarities, and we chose not to downgrade our level of certainty on the basis of this domain.

Imprecision
Our conclusion that caffeine has a beneficial effect using the point estimate does not change when considering the lower and upper level of the 95% confidence interval, because: neither cross the line of no effect; the varying magnitude of effect, from a very small to a small effect, is small; and, a decision to ingest caffeine is unlikely to differ if the true effect is at the upper level of the 95% confidence interval. The latter is because even a very small benefit may make a meaningful difference in sports performance measures such as ranking. Both a statistically and clinically meaningful impact that is ordinarily needed for clinical or public health decisions, is not required. We chose not to downgrade our level of certainty on the basis of this domain.

Publication bias
One quarter of studies had industry involvement, whether through declared interests or funding, which increases the likelihood of publication bias. However, visual and statistical analyses of the data did not provide evidence of publication bias. We chose not to downgrade our level of certainty on the basis of this domain.

Magnitude of effect
Meta-analysis demonstrates that caffeine dose is associated with an observed faster time trial performance when compared to placebo. The magnitude of mean effect is small (-0.26), and at the extremes of the 95% confidence interval (-0.37, -0.16) the effect sizes are small and very small, respectively. We chose not to upgrade our level of certainty on the basis of this domain.

Dose-response gradient
Meta-regression demonstrates that caffeine dose is not associated with time trial

performance (*P*=0.32). Subgroup analysis demonstrates that caffeine dose according to two categories, ≤3 mg/kg BW and >3 to ≤6 mg/kg BW, is not associated with the observed intervention effects on time trial performance. The magnitudes of effect of both groups did not differ significantly (*P*=0.32; Figure 4). We chose not to upgrade our level of certainty on the basis of this domain.

Risk of bias
Five of 39 studies were not randomised and only five studies state a randomisation method. Studies with a short period of wash-out, insufficient for physical recovery, risk carry-over effects in time trial performance in the second or later periods. Identifying the duration of a sufficient wash-out period is a challenge. This is an important potential confounder in the current review, as all 39 studies used a crossover or similar study design. Multiple randomised crossover studies would increase our confidence in the pooled results even if their wash-out periods are insufficient. However, since some studies were not randomised, the risk of carry-over effects remains. One third of studies did not include a familiarisation of the time trial protocol, which may also contribute to carry-over effects. Although a detailed risk of bias assessment was not conducted, we chose to downgrade our level of certainty on this basis.

Indirectness
Duration to complete time trials is a narrow outcome but one directly relevant to the achievement of some sports performance goals (i.e. competitions involving completing a set amount of work in the fastest time and predicted by aerobic exercise performance, in particular cycling time trials which were tested in ~70% of studies) and to a subgroup of athletes who tend to be high caffeine users (i.e., cyclists, rowers, and runners; Aguilar-Navarro et al. 2019). Further, aerobic exercise is a reasonable indirect surrogate outcome of a wide range of other sport types, where aerobic exercise performance is a contributor to success. However, the directness of the evidence is limited when we consider our target population and potential use of caffeine. Most of the pooled sample represent young adult males, who were trained athletes with a high aerobic capacity. This differs to our population of interest for whom food regulation protects, the general Australian and New Zealand population with high prevalence of overweight and obesity, and low levels of physical activity and fitness. We also have a low level of certainty that the small effect observed in the current data will apply to females, other age groups (children, adolescents, and older adults), untrained or unfit people performing sports, or sports people competing in sports where success is not predicted by time trial performance. We chose to downgrade our level of certainty on the basis of this domain.

In deciding whether to downgrade, we assume that the effect of individual variation does not affect the generalisability of our results at a population-level, since the pooled sample is large (n=674) and studies did not exclude volunteers on the basis of individual factors such as genotype. However, there will likely be a subgroup of non-responders for whom this evidence does not translate and caffeine intake has a detrimental impact on their sports performance.

Overall, our certainty in the evidence was downgraded by two levels to a low certainty, due to a risk of bias and indirectness. This means that our confidence in the effect estimate is limited: the true effect may be markedly different from the estimated effect derived from our evidence synthesis.

Discussion
The level of certainty of a body of evidence can be categorised into one of four categories: very low, low, moderate, or high. The certainty of a body of evidence indicates how confident we are that the estimated effect size, from our evidence synthesis, represents the true effect. The definitions are as follows: *very low*, 'We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect'; *low*, 'Our

confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect.'; *moderate*, 'We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different'; and, *high*, 'We are very confident that the true effect lies close to that of the estimate of the effect.' (GRADE Working Group 2013).

Multiple systematic reviews and meta-analyses of caffeine intake on aerobic endurance have been published which include assessments of evidence certainty. An umbrella review by Grgic et al. (2020) identified five systematic reviews (Doherty & Smith 2004, Conger et al. 2011, Ribeiro et al. 2017, Southward et al. 2018, and Shen et al. 2019) with a total of nine meta-analyses, to examine the effects of caffeine on aerobic endurance in healthy individuals. The level of certainty of the body of evidence included in each meta-analysis was categorised by Grgic et al. (2020) as low for four meta-analyses, very low for two, and moderate for three. In a separate publication, meta-analytical evidence to evaluate the effect of caffeine on aerobic endurance performance in soccer players was graded by a different pair of reviewers. They graded the body of evidence as having a very low (four meta-analyses) to low (one) level of certainty (Ferreira et al. 2021). The level of certainty in the evidence that FSANZ assessed is therefore consistent with that of 14 published meta-analyses.

Table 2: Summary of included studies: study design and sample characteristics

First author (year)	Study design	Randomisation (Y/N/Other) and randomisation method	Blinding (D/S, who, efficacy)	n	Training / fitness level[2]		Sex (n)	Habitual caffeine intake	Location	Industry involvement	Selective reporting
					Publication	VO_{2max} or VO_{2peak}					
Acker-Hewitt (2012)	Counterbalanced	N. Described as semi-randomised. The initial treatment trial was always the placebo trial, and authors imply there was randomisation among the following three intervention trials.	D	10	ns	T	M	ns	USA	ns	Two (from an original 12) subjects withdrew prior to completion because of "circumstances unrelated to the investigation."
Astorino (2011) and Astorino (2012a) Endurance-trained subgroup[1]	Crossover	Treatment order was assigned using a Latin Squares design.	S Participants. "Only 5 (31%) were able to differentiate between the treatments."	8	Endurance-trained (≥5 h/week; competing in sports including cycling, running or triathlon).	T	M	2-7 d/week; range: 15-320 mg/d and none >350 mg/day (67.5 ± 53.9 mg/d)	USA	N	
Astorino (2011) and Astorino (2012a) Recreationally active subgroup[1]	Crossover	aa	aa	8	Recreationally active (≥5 h/week; participating in team sports, resistance training, and/or cardiovascular exercise).	U	aa	2-7 d/week; range: 15-320 mg/d and none >350 mg/day (125.6 ± 120.6 mg/d)	aa	aa	
Astorino (2012b)	Crossover	Y (method: ns). Treatment order was assigned using a Latin Squares design.	S Participants. "One subject (11%) was able to differentiate between the treatments."	9	Competitive in cycling or triathlon. Training ≥5 h/week.	T	M (8) & F (1)	2-7 d/week and none >300 mg/day (65.6 ± 50.8 mg/d)	USA	N	
Astorino (2012c)	Crossover	Y (method: ns). Treatment order was	D Co-investigators and	10	Active (≥3 d/week; resistance training,	NC	F	Range: 0-7 d/week. Mean	USA	N	

		assigned using a Latin Squares design.	participants (not primary investigator). "Subjects were unable to distinguish between drinks across days"		aerobic exercise and/or recreational sports but minimal cycling experience; current physical activity: 6.7 ± 4.0 h/week)			intake = 151.1 ± 107.4 mg/d. One participant not a habitual caffeine consumer.			
Bell (2002)	Crossover	Y (method: ns).	D	12	Recreational runners	T	M (10) & F (2)	Six were regular coffee drinkers, i.e., >1 cup/d. Six were irregular or non-caffeine users, i.e., <3 cups/week.	CAN	Y Sandoz Canada provided caffeine and Roberts Pharmaceutical Canada provided ephedrine.	
Bortolotti (2014)	Crossover	Y (method: ns).	D	13	Cyclists with a minimum of 2 y competitive cycling experience, 253 ± 142 km/week, and free of injuries for minimum of 6 mo.	NC	M	ns	BR	N	
Carr (2011)	Crossover, semi-counterbalanced	N. No mention of randomisation or allocation. Only states it was administered in a semi-counterbalanced fashion and all subjects completed two baseline and four experimental trials. Two cohorts (n=6 at Perth and n=2 at Canberra, Australia).	D Six subjects correctly noted when they had been given the caffeine treatment and 5 when they had been given the placebo.	8	Well-trained rowers who had competed at the Australian Rowing Championships and of which 6 were scholarship holders at the Western Australian Institute of Sport. Personal-best 2000 m ergometer times of 6:24.6 min:s ± 12.9 s for men and 6:57.0 ± 2.1 s for women.	NC	M (6) & F (2)	ns	AUS	N	
Cohen (1996)	Crossover	Y (method: ns).	D "Each racer was able to	7	Endurance trained and heat	NC	M (5)	Range of 0-300 mg/d. None of	USA	Y Supported in	

			identify correctly that they had ingested caffeine when they were on the 9 mg/kg BW dose."		acclimatised competitive road racers (best marathon time: 172.56 ± 22.32 min), running a minimum of 64 km/week (82.13 ± 24.69 km/week), with the last 3 mo training in hot and humid weather.		& F (2)	the subjects used caffeine as an ergogenic aid to their racing.		part by a Gatorade Sports Science Institute Student Research Award.	
Conway (2003)	Factorial	Y (method: ns). Balanced factorial trial.	D	8	Well-trained cyclists and triathletes.	T	M	None of the subjects reported habitual consumption of large amounts of caffeine (<250 mg caffeine/d)	AUS	N	One (from an original 9) subject was excluded from analysis because he was unable to complete the time trial when given the single dose of caffeine as he experienced nausea and nervousness.
Cox (2002) Study A	Latin-square	ns	S (only for the caffeine and placebo condition; not for the Coca-Cola condition which is not used in the current review). Participants. ". . . 5 of the 12 subjects were able to correctly identify the order of treatments in the Precaf, Durcaf, and Placebo trials."	12	Highly trained cyclists and triathletes, cycling >250 km/week.	T	M	Subjects' background caffeine intake ranged from occasional intake during competitive events to habitual daily intake of ~150 mg/d.	AUS	Y Grant from Nestle Australia to the Department of Sports Nutrition at the Australian Institute of Sport.	
Desbrow (2009)	Incomplete Latin-square	Y (method: ns).	D ". . .two subjects who were able to correctly identify the order of their caffeine doses; however, only one	9	Trained cyclists or triathletes.	T	M	Subjects' habitual average caffeine intake estimated as 232 ± 129 mg/d (range: 74 to 395 mg/d).	AUS	Y Funded by a Sports Dietitians Australia/Gatorade student	

			of these subjects indicated a high degree of certainty over their predictions."							research grant. Nestle Australia covered all food expenses.	
Desbrow (2012)	Crossover	Y (method: ns).	D "Six of the 16 participants (38%) correctly identified the treatment order of all three trials and five participants (31%) could distinguish when they had received the placebo treatment, but were unable to differentiate between the two caffeine trials. The majority (n=12; 75%) believed that they performed better when caffeine was ingested . . ."	16	Well-trained cyclists.	T	M	210 ± 115 mg/d, range 10–600 mg/d.	AUS	N	
Felippe (2018)	Counterbalanced	Y (method: ns).	D	11	Moderately trained cyclists (258 ± 24 km/week) with regular participation in local competitions (~11 competitions/year).	T	M	ns	BR	N	
Glaister (2015)	Counterbalanced, Latin-square	Y (method: ns).	D	14	Competitive cyclists and triathletes (training 10.7 ± 2.2 h/week).	NC	F	210 ± 131 mg/d	UK	Y Bionox Ltd. provided help with plasma nitrate/nitrate analyses.	
Gonçalves (2017)	Crossover, counterbalanced, Latin-square	Y (method: ns). "randomly assigned"	D “17 participants correctly guessed the supplement ingested. Moreover, 13 and 17 participants did	40	Endurance-trained cyclists (>150 km/week).	U	M	Across tertiles (mg/d): 58 ± 29 (n=14), 143 ± 25 (n=12), 351 ± 139 (n=14)	BR	N	Two (from an original 42) were excluded from the analysis because one did not complete the habitual caffeine

			not know what supplement they had ingested during caffeine and placebo. In addition, 12 and 8 participants incorrectly guessed the supplement ingested during caffeine and placebo. Fisher's exact test did not show any significant differences among trials for the proportion of supplement identification (*P*=0.57)."								intake form and another did not complete all exercise trials for reasons unrelated to the study.
Graham-Paulson (2016)	Repeated measures	ns	D Two participants correctly identified the treatment in all four trials.	11	Recreationally-trained	U	M	160 ± 168 mg/d	UK	N	
Guest (2018)	Split-plot	Y. Method: randomisation of treatment order was "done using balanced permutations blocked by time of entry (randomization.com)."	D 31% of the caffeine trials were correctly identified as containing caffeine (81% were identified incorrectly as not containing caffeine, and 19% were identified as "maybe caffeine". Only 3% of subjects correctly identified all three trials i.e. 2 caffeine, 1 placebo.)	101	Competitive athletes from endurance (42%), power (42%), or mixed (16%) sports. Training or competing at least 8 h/week, for 9 out of 12 mo.	U	M	Across three genotypes (AA, AC, CC): dietary caffeine (mg/d; excludes caffeine for sports performance) 87 ± 18, 80 ± 20, 38 ± 24; caffeine for sports performance (mg/d) 61 ± 13, 89 ± 17, 80 ± 74.	CAN	Y Funded by the Canadian Institutes of Health Research, the Canadian Foundation for Dietetic Research, Nutrigenomix Inc., The Coca-Cola Company, and Mitacs. Two authors report industry association: A.E.-S. is the Founder and holds shares in	Twelve (from an original 113) subjects were excluded from analysis because eight dropped out of the study due to a sport-related injury (n=3), school or work demands (n=2), unwillingness to abstain from caffeine (n=2), or relocation (n=1); and an additional four subjects were excluded because of incomplete data.

										Nutrigenomix Inc., and N.G. serves on the Scientific Advisory Board of Nutrigenomix Inc.	
Hodgson (2013)	Crossover, counterbalanced	Y (method: ns).	S Participants. Six of the 8 subjects correctly guessed the caffeine arm.	8	Trained cyclists or triathletes (training 3 or more per week, with >90 min /session).	T	M	≤300 mg/d.	UK	Y One author (A.E.J.) was employed by Pepsi Co.	
Hulston & Jeukendrup (2008)	Crossover	Y (method: ns). "trials were performed in random order"	D	10	Endurance trained cyclists	T	M	186 ± 101 mg/d, range: 70-400 mg/d.	UK	Y Glaxo SmithKline Consumer Healthcare, United Kingdom.	
Hunter (2002)	Crossover, repeated measures	Y (method: ns).	S Participants Seven of the 8 subjects correctly identified when they had ingested caffeine.	8	Competitive, endurance-trained cyclists, cycling 200-500 km/week.	T	M	ns	ZA	N	Seven of the original 15 subjects were excluded because they were unable to achieve the required cycling speed (n=2) or because they found the trial “too arduous” (n=5).
Jacobson (2001)	Crossover	Y (method: ns). Trials were "performed in random order"	D	8	Endurance-trained cyclists and triathletes (mean cycling 366 km/week).	T	M	Subjects were not habitual caffeine users.	AUS	N	
Kovacs (1998)	Crossover	Y (method: ns).	D	15	Trained triathletes and cyclists (≥2 h/day and ≥4 times/week).	NC	M	Range: 20-410 mg/d.	NL	Y Grant from Novartis Nutrition, Ltd., Bern, Switzerland.	One (from an original 15) was excluded from statistical analysis because caffeine had been detected in plasma during a placebo trial.
MacIntosh	Crossover	Conditions were	D	11	Competent	NC	M	<300 mg/week	CAN	N	

(1995)		"assigned in a random . . . manner"			distance swimmers (<25 min for 1500 m).		(7) & F (4)				
Miller (2014)	Crossover, counterbalanced	Randomised (ns).	D "After the first and second trials, 50 and 66% of participants were able to correctly identify their treatment, respectively."	6	Well trained cyclists and triathletes (cycling ≥250 km/week).	T	M	< 50 mg/d (n=5) and ~300 mg/d (n=1).	AUS	N	Four (from an original 10) participants discontinued after the familiarisation session secondary to work constraints and injury unrelated to this study.
O'Rourke (2008) Well-trained subgroup[1]	Crossover	Conditions were "randomly administered"	D	15	Well trained runners (at least 5 years club level competition experience).	NC	ns	None of the participants was considered to be a habitual caffeine consumer.	AUS	ns	
O'Rourke (2008) Recreationally active subgroup[1]	Crossover	aa	aa	15	Recreational runners (most with a history of playing team-sports, such as hockey).	NC	aa	aa	aa	aa	
Pitchford (2014)	Crossover, counterbalanced	Randomised (ns).	D Five participants correctly identified the trial they consumed caffeine pre-exercise.	9	Highly trained cyclists	T	M	Seven of the nine participants were regular caffeine consumers in the average range of 100–300 mg/d.	AUS	N	
Potgieter (2018)	Crossover	Randomised. "An independent laboratory (African Micronutrient Research Group) randomized . . . groups . . ."	D	26	Trained triathletes (12.8 ± 4.5 h/week)	NC	M (14) & F (12)	413 ± 505 mg/d	ZA	N	
Quinlivan (2015)	Crossover, incomplete Latin-square	Y. Method: Conditions were "randomly administered" and "randomized by a person independent	D Participants and investigators. "No participant reported certainty about all treatment being	11	Trained cyclists and triathletes	T	M	271 ± 29.5 mg/d	AUS	N	One (from an original 12) was forced to withdraw after the initial trial due to severe stomach pains experienced after

		to the study using an incomplete Latin-square design"	administered. Only 1 participant correctly determined the order of all treatments. Three participants correctly identified the trial in which they received Red Bull; however, only 1 participant was certain of this choice. Five participants believed they received Red Bull after being administered the alternative drink (3 during the placebo trial, 2 during caffeine trial). Eight participants thought they were given an alternative caffeinated energy drink (not Red Bull) during all 3 trials)." (Note: the Red Bull arm is not used in the current review.)								consuming Red Bull (a third condition, not reported here).
Roelands (2011)	Crossover	Y (method: ns).	D	8	Trained cyclists and triathletes	NC	M	108 ± 47 mg/d	BE	N	
Santos (2013)	Repeated measures, crossover	Y (method: ns).	D	8	Recreationally trained cyclists (~223 km/week).	T	M	ns	UK	N	
Scott (2015)	Counterbalanced, repeated measures, crossover	ns	S	13	Trained in a mixture of sports (who competed in the British Universities and Colleges Sport competition).	NC	M	82 ± 59 mg/d	UK	Y Gels were provided by Science in Sport, Blackburn, UK.	
Skinner (2010)	Crossover	Y. Method: "The order of trials was randomized by a person independent	D "Five subjects correctly identified each of the placebo and 6 mg/kg,	10	Competitive rowers (national-level)	U	M	<400 mg/d	AUS	N	

		of the project using a random number-generating process."	whereas only one subject correctly identified each of the 2 and 4 mg/kg trials. Of these, only three subjects correctly identified two of the four trials, whereas six correctly identified one trial only.")								
Skinner (2013)	Crossover	Y. Method: "The trial order was randomised by a person independent to the project using a random number generating process."	D	14	Trained cyclists and triathletes	T	M	ns	AUS	N	
Spence (2013)	Repeated measures	Y (method: ns).	D "3 out of 10 correctly guessed when they were using caffeine, and 4 out of 10 correctly guessed when they were given the PLA or the PSE." (Note: the PSE arm is not used in the current review.)	10	Trained cyclists and triathletes	T	M	ns	AUS	N	
Stadheim (2013)	Crossover	Y (method: ns).	D ". . .subjects were unable to sense which product they received during the different trials."	10	Highly trained cross-country skiers (who compete in the Norwegian National Cross country Skiing Cup).	T	M	Four subjects were regular caffeine drinkers (100–250 mg/d). One subject normally had a high daily intake of CAF (>300 mg/d). The remaining subjects' intakes are not reported.	NO	N	
Stadheim (2015)	Crossover	Y (method: ns).	D ". . .subjects were unable to sense which product	13	Subelite cross-country skiers (who compete in the	T	M	ns	NO	N	

			they received during the different trials."		Norwegian National Cross country Skiing Cup).						
van Nieuwenhoven (2005)	Crossover	Y (method: ns).	ns	98	Well trained	NC	M (90) & F (8)	ns	NL	N	
Wemple (1997)	Counterbalanced	Y (method: ns). "randomly assigned"	D Subjects reported no awareness of caffeine treatment during the trials.	6	Highly active (aerobic exercise 5+ d/week, resistance training 2+ d/week).	U	M (4) & F (2)	Prior to the study three subjects regularly consumed 2-3 cups of coffee each day (~300 mg/d), whereas the other three abstained from caffeine.	USA	Y Partially funded by a student research grant from the Gatorade Company. Chicago. IL.	
Womack (2012) AA homozygotes subgroup	ns but appears to be crossover	Y (method: ns). "randomly administered"	D	16	Recreationally competitive cyclists	T	M	85.71 ± 106.49 mg/d (n=16)	USA	N	One (of an original 36) participant was excluded from the study post-hoc, as their cycling performance differed by more than two standard deviations from the mean value of the group.
Womack (2012) C allele carriers subgroup	aa	aa	aa	19	aa	T	aa	86.62 ± 145.40 mg/d (n=19)	aa	aa	One participant excluded (aa).

ns, not specified (i.e. no further details provided by publication); aa, as above; Y, yes; N, no; n, sample size (unit: individuals); S, single-blinded; D, double-blinded; h, hour; d, day; mo, month; y, year; T, trained; U, untrained; NC, not categorised; M, male; F, female; AUS, Australia; BE, Belgium; BR, Brazil; CAN, Canada; NO, Norway; NL, The Netherlands; UK, United Kingdom; USA, United States of America; ZA, South Africa.

[1] Description of trained versus untrained represents the investigators' categorisation. Our categorisation is the same for Astorino (2011 and 2012a) but different for O'Rourke (2008) which we classed as 'uncategorised' due to no VO_{2max} and VO_{2peak} being reported.

[2] First column represents the investigators' categorisation described in the publication. Second column represents our categorisation based on a threshold, ≥55 mL/kg/min VO_{2max} or VO_{2peak}.

Table 3: Summary of included studies: diet and exercise, and intervention and comparator characteristics

First author (year)	Diet and exercise prior to, and on the day of, trial (intervention and comparator conditions)	Placebo characteristics (comparator condition)	Caffeine characteristics (intervention condition)			
			Dose (mg/kg BW)	Single or split dose	Caffeine and non-caffeine/placebo (where provided) characteristics	Timing of caffeine[2]
Acker-Hewitt (2012)	Maintain consistent dietary habits within 48 h of TT. No caffeine and alcohol for 24 h prior to each TT. Participants were instructed to eat a self-selected meal no less than 12 h prior to the start of each trial (i.e., dinner on the evening prior to testing). Maintain training habits, except no heavy exercise within 48 h of TT. On TT day: fed standardised meal (breakfast, 2 h prior to TT, one from three choices). Artificially sweetened water (250 mL) was administered at the following time points during the exercise trial: immediately prior to exercise, following the 20-min SS, and 20 min into the TT.	Pill	6	Single	Anhydrous caffeine, commercially available (TerraVita), pill.	60 min prior to exercise (80 min prior to TT)
Astorino (2011) and Astorino (2012a) Endurance-trained subgroup[1]	Subjects asked to "follow the same diet on the day before each trial". No caffeine within 48 h before TT. Maintain exercise volume and intensity, except no intense lower body exercise for 48 h prior to TT. On TT day: subjects were "3 hours post-absorptive".	Beverage: one package of commercially available, noncaloric, lemon-flavored beverage (Crystal Light, Northfield, IL), 5 mg/kg BW of glucose, and 125 mL of noncaloric carbonated soda. Participants mixed each drink with 250 mL of cold water.	5	Single	Anhydrous caffeine, (Gallipot, St. Paul, MN), added to beverage (same as C).	1 h prior to exercise (70 min prior to TT)
Astorino (2011) and Astorino (2012a) Recreationally active subgroup[1]	Diet: aa. Exercise: aa. On TT day: aa.	aa	5	aa	aa	aa
Astorino (2012b)	Diet: aa. Exercise: aa. On TT day: aa.	aa	5	Single	aa	aa
Astorino (2012c)	Subjects asked to "follow the same diet on the day before each trial". No caffeine within 24 h of TT. No lower body exercise within 24 h of TT.	Beverage: 113 mL of diet 7-up, 113 mL of water, and 6 mg/kg BW of glucose.	6	Single	aa	aa

	On TT day: ns.					
Bell (2002)	No alcohol for 24 h and caffeine before (duration of time: ns) TT. Exercise: ns On TT day: fasted for 8-12 h. Then, 20 to 30 min after placebo or caffeine ingestion, a light meal consisting of juice and a muffin was eaten (nutrient composition: ns).	Capsules, 300 mg of a dietary fibre (Metamucil®, Procter & Gamble, Toronto, ON).	4	Single	Capsule (Sandoz, Canada)	1.5 h prior to TT (no warm up)
Bortolotti (2014)	No caffeine 48 h before TT. No alcohol 24 h before TT. No strenuous exercise 24 h before TT. On TT day: ns.	Maltodextrin	6	Single	Capsule	50 min prior to exercise (1 h prior to TT)
Carr (2011)	Maintain same diet for 24 h before every TT. No caffeine for 48 h before each TT. Maintain the same training pattern (type, duration, intensity) for the 24 h before every TT. On TT day: fasted (overnight fast).	Dose 1 (I+C): 90 min before warm up exercise (97 min before TT), participants consumed placebo capsules containing corn flour (White Wings foods, NSW, Australia). Dose 2 (C only): 30 min before warm up exercise (37 min before TT, participants consumed glucose (Glucodin, NSW, Australia) capsules.	6	Single	Dose 1: same as C. Dose 2: No Doz (Key Pharmaceuticals Pty. Ltd., NSW, Australia) in gelatin capsules. (We note: according to website: No Doz contains Vitamin B1 and B3, in a base containing glucose of an unknown quantity.)	30 min before warm up exercise (37 min before TT)
Cohen (1996)	Maintain normal diet during the experiment. No caffeine and alcohol for 24 h prior to TT. Maintain normal training regime, except no vigorous exercise for 24 h prior to TT. On TT day: fasted for 9 h. However, later the authors state "To mimic a typical race as closely as possible, the subjects were encouraged to perform their usual prerace rituals and dietary habits . . .". Note: this assessment has not categorised this study as having subjects who consumed a meal, snack or liquid meal within 3 h of the experiment.	Baking flour, capsule.	5 or 9	Single	Anhydrous caffeine, capsules.	1 h prior to TT (no warm up)
Conway (2003)	No caffeine for 48 h before TT. Repeat the same food selection before each TT. Repeat the same training before each TT. Subjects were asked to refrain from heavy exercise 24 h before TT.	Lactose, gelatin capsule.	6	Single or split (two half doses)	Gelatin capsule.	I (Single): 60 min prior to exercise start (155 min prior to TT).

	On TT day: fasted for 12 h.					I (Split): caffeine (half dose) 60 min prior to exercise start (155 min prior to TT) and caffeine (half dose) 45 min after start of exercise (50 min prior to TT).
Cox (2002) Study A	No caffeine for 48 h before each TT. Food intake was standardised for 24-48 h before each TT. During the 24 h immediately before each TT, subjects were provided with a prepacked standard diet (200 kJ/kg BW, 63% CHO or 8 g CHO/kg BW, 20% fat, and 17% protein). Training was standardised for 24-48 h before each TT. No training for 24 h before TT. On TT day: fasted for 12-14 h. Then fed a standardised CHO-rich meal providing 2 g CHO/kg BW (a standard breakfast of fruit juice, toasted bread and jam, and a Power Bar). Exercise commenced 2 h after intake of the meal.	Capsule (polycose) and beverage (commercial sports drink: 6.3% CHO, 18 mmol/L sodium; providing 7 x 5 mL/kg with a total CHO of 2.1 g/kg BW over ~2.5 h during exercise (including 120 min of SS cycling followed by TT)).	6	Single	Capsule. Beverage (same as C).	1 h prior to exercise (3 h prior to TT).
Desbrow (2009)	No caffeine for 24 h before each TT. During the 24 h immediately before each TT, subjects were provided with a prepacked standard diet (200 kJ/kg BW, 64% CHO or 8 g CHO/kg BW, 24% fat, and 12% protein). No heavy training 24 h before each TT. Any light training was to be completed by 1200 h the day before TT. On TT day: in the morning, 2 h prior to experiment, subjects consumed a standardised and pre-prepared meal that "provided 2 g CHO/kg BW and included a 600 mL commercial sports drink (Gatorade), fruit bread, jam, and a PowerBar)".	Metamucil®, capsule. Beverage (6% CHO solution; 8 mL/kg BW at t=0 min of SS cycling; 5 mL/kg BW at t=20 min, 40 min, 60 min, 80 min, 100 min, and 120 min of SS cycling; and, ~5 mL/kg BW during TT.).	1.5 or 3	Single	Caffeine citrate (PCCA, Houston, TX) capsule. Beverage (same as C).	1 h prior to exercise (3 h prior to TT).
Desbrow (2012)	No caffeine and alcohol for 24 h before each TT. For the 24 h period immediately before each TT, subjects were provided with a prepacked standardised diet (200 kJ/kg BW, 7.5 g CHO/kg BW). No strenuous exercise for 24 hr prior to TT. On TT day: in the morning, subjects consumed a light pre-exercise meal (42 kJ/kg BW, including 2 g CHO/kg BW).	400 mg Metamucil, 100% psyllium husk fibre, capsule. Beverage (6% carbohydrate-electrolyte; 3 mL/kg BW during the warm-up, and upon completion of 30% and 60% of the target amount of work).	3 or 6	Single	Anhydrous caffeine, capsule. Beverage (same as C).	~90 min prior to exercise/TT.
Felippe (2018)	No alcohol and caffeinated beverages for 24 h before each visit.	Cellulose, gelatin capsule.	5	Single	Gelatin capsule.	Authors state "~75

	Maintain same diet for the 24 h before each TT. No vigorous physical activities 24 h before each TT. On TT day: participants consumed their last meal 2 h before each test session.					min prior to exercise" but according to the protocol, it appears to be ~57 min prior to TT.
Glaister (2015)	Maintain normal diet during the testing period, and repeat the same diet for 24 h before each TT. No caffeine- and nitrate-rich foods for 24 and 48 h, respectively, before each TT. No strenuous exercise for 24 h before each TT. On TT day: subjects were instructed to avoid food and drink in the hour before each TT.	Maltodextrin (My Protein, Manchester, United Kingdom), gelatin capsule. Beverage, 70 mL dose of concentrated beetroot juice (Beet IT Sport Shot; James White Drinks, Ltd., Suffolk, United Kingdom) with the nitrate content removed (placebo: ~0.01 nmol nitrate). (We note: according to website, a 70mL shot provides 12.6 g CHO).	5	Single	Gelatin capsule (Sigma-Aldrich, Steinheim, Germany). Beverage (same as C).	1 h prior to trial (70 min prior to TT, if 'trial' means start of exercise).
Gonçalves (2017)	Subjects instructed to abstain from alcohol, and caffeine-containing substances within the 24-h period before TT (they were provided with a comprehensive list of the main products containing caffeine). No training 24 h prior to TT. On TT day: fasted for 6 h.	Dextrose, gelatin capsule.	6	Single	Anhydrous caffeine, gelatin capsule.	1 h prior to trial (65 min prior to TT, if 'trial' means start of exercise).
Graham-Paulson (2016)	No caffeine and alcohol within the 24 h before each TT. Repeat the same diet for 24 h before each TT. No exercise 24 h prior to each TT. On TT day: participants consumed a self-selected standardised meal 1.5 h prior to arriving at the laboratory, which was noted upon arrival (62 ± 10% CHO, 18 ± 9% protein, 20 ± 9% fat).	Dextrose (Bulk Powders, Colchester, UK), capsule.	4	Single	Anhydrous caffeine (Bulk Powders, Colchester, UK), capsule.	45 min prior to exercise (75 min prior to TT).
Guest (2018)	Maintain regular diet and sleeping habits, and abstain from caffeine 1 week before the first visit and for the duration of the data collection (4 weeks total). Repeat the same diet for the 24 h before each TT. No strenuous activity 48 h before each visit. On TT day: ns.	Dextrose, capsule.	2 or 4	Single	Anhydrous caffeine (A&C American Chemicals Ltd., Saint-Laurent, Quebec, Canada), capsule.	25 min before exercise (~40 min before TT).
Hodgson (2013)	Repeat the same diet for the 24 h before each TT, as well as avoid alcohol and caffeinated products.	8 mg quinine sulphate (Sigma, UK) dissolved in beverage (600 mL	5	Single	Anhydrous caffeine (99.8% pure; Blackburn Distributions	1 h prior to exercise (1.5 h prior to TT).

	No exercise in the 24 h prior to TT. On TT day: fasted for 8 h, then provided placebo or caffeine.	water).			Ltd, Nelson, United Kingdom), dissolved in beverage (600 mL water).	
Hulston & Jeukendrup (2008)	Repeat the same diet before each TT (time period is not defined). No avoid alcohol and caffeine intake for 24 h before TT. Subjects also asked to follow a specific exercise/diet regimen starting 5-7 d before each TT. No strenuous exercise for 24 h before TT. On TT day: fasted for 10-12 h.	Beverage, 6.4% glucose solution.	5.3	Split (> 2 doses)	Beverage, 6.4% glucose plus caffeine.	During warm up: 5.5 mL/kg at the onset of exercise, then 2 mL/kg every 15 min during SS.
Hunter (2002)	No caffeine for 48 h before each TT. The day prior to TT, subjects followed a prescribed diet, which consisted of 60% CHO or 5 g CHO/kg BW and 17% protein or 1.3 g protein/kg BW. Subjects repeated the same dietary regimen before each TT. Maintain same type of training for the duration of the trial and no heavy physical exercise on the day before TT. On TT day: 3 h before TT, subjects consumed a standardised breakfast (30 g of cornflakes and 150 mL of 2% fat milk).	Pre-TT: gelatin capsules containing white flour with 150 mL of a sports electrolyte solution containing 7% CHO. During TT: gelatin capsules (flour) and 150 mL drink (as above) every 15 min until trial completion.	9	Split (> 2 doses)	Pre-TT: gelatin capsules containing caffeine (6 mg/kg BW) with 150 mL of a sports electrolyte solution containing 7% CHO. During TT: caffeine (0.33 mg/kg BW) and 150 mL drink (as above) every 15 min until trial completion. (Total caffeine = 9 mg/kg BW over ~2.5 h.)	Following isometric testing, subjects ingested capsules/beverage (1 h before warm up) followed by staggered doses during the TT.
Jacobson (2001)	Subjects refrained from consuming caffeine (coffee and soft drinks) for 72 h prior to TT. In the 24 h prior to TT, subjects consumed a prepacked standard diet (0.21 J/kg BW, 63% CHO or 8 g CHO/kg BW, 20% fat and 17% protein). 24 h prior to TT, subjects completed a standardised 60 min cycling bout at a work rate equivalent to ~70% VO_{2max}. No further strenuous physical activity. On TT day: fasted for 12-14 h. Subjects consumed a meal including 2.6 g/kg BW of high glycaemic index CHO (glucose polymer Polyjoule, Sustagen Sport powder, chocolate flavouring and skim milk), immediately followed by caffeine or placebo.	~500 mg sucrose, capsule.	6	Single	Capsule.	55 min prior to exercise (180 min prior to TT).
Kovacs (1998)	Repeat the same diet and activities protocol before all remaining tests (text implies 2 days before TT, but it is unclear). No caffeine 48 h prior to TT. No exhaustive training 48 h prior to TT.	Beverage (water with 68.8 g CHO/L and electrolytes; 14 mL/kg BW volume total).	2.1, 3.2, or 4.5	Split (> 2 doses)	Beverage (same as C plus caffeine). Caffeine was one of three doses: 150, 225, or 320 mg (equating to a mean intake 2.1, 3.2, or 4.5 mg/kg BW, respectively).	8 mL/kg ingested at same time as start of first warm up (20 min warm up). 3 mL/kg of beverage ingested twice

	On TT day: fasted overnight. Subjects consumed a standardised breakfast (1.5 g/kg BW of bread, 0.5 g/kg BW of cheese, 10 g of butter, and 200 mL of mineral water) ~30 min prior to start of exercise.					during TT (at t=20 and 40min of TT).
MacIntosh (1995)	No caffeine 48 h prior to TT. No strenuous exercise 24 h prior to TT. On TT day: fasted for ≥6 h.	Beverage: 200 mL of artificially sweetened fruit drink.	6	Single	Beverage (same as C plus caffeine citrate and extra artificial sweetener).	2 h prior to exercise (2.5 h prior to TT).
Miller (2014)	No caffeine and alcohol 24 h prior to TT. Participants were prescribed a standardised diet (CHO: 7 g/kg/day) for the 24 h prior to TT. No strenuous exercise 24 h prior to TT. On TT day: fasted for 4 h.	Lactose (gelatin capsule). Beverage (3 mL/kg BW every 15 min during training session; 7.4 g CHO/100 mL and electrolyte drink, Powerade) providing ~1 g CHO/kg BW/h.	6	Split (two half doses)	Gelatin capsule. Beverage (same as C).	First 3 mg/kg BW 1 h before, and second 3 mg/kg BW into, the training session.
O'Rourke (2008) Well-trained subgroup[1]	24 h prior to TT, participants consumed ~8-10 g CHO/kg BW and enough fluid to ensure urine was clear in the hours before the TT. No caffeine 48 h prior to TT. No strenuous exercise 48 h prior to TT. On TT day: ns.	Sugar formulated tablet.	5	Single	Tablet. No Doz (Key Pharmaceuticals Pty. Ltd., NSW, Australia) containing ~100 mg caffeine/tablet. (We note: according to website: No Doz contains Vitamin B1 and B3, in a base containing glucose of an unknown quantity.)	1 h prior to TT.
O'Rourke (2008) Recreationally active subgroup[1]	Diet: aa. Exercise: aa. On TT day: ns.	aa	5	aa	aa	aa
Pitchford (2014)	Participants were provided with a pre-packaged diet, consisting of all food and fluid for the day preceding the TT (200 kJ/kg BW and 7.5 g CHO/kg BW) and a breakfast meal for the morning of the TT (40 kJ/kg BW and 1.5 g CHO/kg BW). No alcohol for at least 24 h and caffeine for at least 12 h prior to each TT. No physical activity aside from activities of daily living for 24 h before TT. On TT day: subjects consumed a standardised breakfast meal.	Psyllium husk (Metamucil®, P&G Australia Pty Ltd., Sydney, NSW, Australia). Beverage (3 mL/kg BW of a carbohydrate-electrolyte beverage, Gatorade at every 25% of TT completion).	3	Single	Anhydrous caffeine (PCCA, NSW, Australia), capsule. Beverage (same as C).	90 min prior to TT.
Potgieter (2018)	No caffeine for 14 d prior to TT.	Artificial sweetener (capsule).	6	Single	70% w/w caffeine (Maxx Performance Inc., Roanoke,	60 min prior to TT.

	No competition racing for 14 d prior to TT. No exhaustive exercise for 48 hr prior to TT. On TT day: on arrival subjects were fasted and then consumed a self-supplied pre-event meal. Pre-race meal or exercise regimes were not prescribed, but participants had to duplicate preparations for the two racing days.				VA), capsule.	
Quinlivan (2015)	No alcohol and caffeine 24 h prior to TT. Participants received a standardised pre-packaged diet to follow for 24 h before TT (200 kJ/kg BW and 7.5 g CHO/kg BW). No strenuous exercise 24 h prior to TT. On TT day: fasted overnight. 90 min prior to TT, participants consumed a pre-trial meal (1.5 g/kg BW of Powerbar), which in combination with the drink ensured participants ingested 42 kJ/kg BW with ~2 g CHO/kg BW before the TT. During warm-up, and at 30% and 60% of TT completion, participants were provided with ~3 mL/kg BW of commercial sports drink (Gatorade).	Capsule (Metamucil™, P&G Australia Pty Ltd., Sydney, NSW, Australia). Beverage (9.4 mL/kg BW; Gatorade powder mixed with carbonated water, containing 185 kJ/100 mL and 11 g CHO/100mL).	3	Single	Anhydrous caffeine (PCCA, USA), capsule. Beverage (same as C).	90 min prior to TT.
Roelands (2011)	Maintain the same diet and refrain from caffeine for the 2 d prior to each TT. No alcohol for 24 h before TT. Maintain the same physical activity for the 2 d prior to each TT, with no exercise in the last 24 h prior to TT. On TT day: Participants consumed a standardised breakfast (90-100 g CHO).	Lactose (capsule).	6	Single	Capsule.	1 h prior to exercise (2 h prior to TT).
Santos (2013)	Maintain the same diet and refrain from caffeine for the 24 h prior to each TT. No heavy training in the 24 h prior to TT. On TT day: fed. Participants consumed a standardised breakfast which consisted of 60% CHO, 25% lipids and 15% protein.	Cellulose (capsule).	5	Single	Capsule.	1 h prior to exercise (1 h 10 min prior to TT).
Scott (2015)	Maintain the same diet and refrain from alcohol and caffeine for 24 h prior to each TT. Maintain usual exercise pattern, but no strenuous exercise 24 h prior to TT. On TT day: fasted for 12 h.	Go Isotonic Energy Gel (60 mL), Science in Sport, energy 367.2 kJ, CHO 21.6 g.	100 mg (absolute not relative dose; equates to 1.3 mg/kg BW, range:	Single	Smart 1 Energizer Gel (60 mL), Science in Sport, energy 367.2 kJ, CHO 21.6 g, caffeine 100 mg.	10 min prior to TT.

			0.98-1.47 mg/kg BW)			
Skinner (2010)	Participants consumed a standardised diet of 200 kJ/kg BW and 8 g CHO/kg BW which excluded caffeine (24 h prior to the TT) and included a pre-TT meal of 2 g CHO/kg BW (consumed 90 min prior to TT). Maintain normal training regimen, but no strenuous exercise 24 h prior to TT. On TT day: ns.	Calcium sulfate, capsule.	2, 4 or 6	Single	Anhydrous caffeine, capsule.	60 min prior to TT (40 min prior to exercise).
Skinner (2013)	Participants consumed a high CHO meal the night before TT. For 48 h pre-TT: no from caffeine and substances which potentially affect the metabolism of caffeine (e.g. all cruciferous vegetables, charcoal-broiled beef, aspirin, cimetidine), alcohol. For 48 h pre-TT: maintain a hydrated state. Repeat the same food intake during the 24 h pre-TT. No vigorous physical activity for 24 h prior to TT. Light training permitted until midday the day prior to TT. On TT day: fasted for ≥12 h. Pre-exercise meal (51 ± 3 kJ/kg BW including 2 ± 0 g CHO/kg BW) was consumed 20 min prior to ingestion of capsules (i.e. 80 min prior to TT).	Calcium sulfate, gelatin capsule.	6	Single	Anhydrous caffeine (Sigma-Aldrich), gelatin capsule.	1 h prior to TT.
Spence (2013)	No caffeine 48 h prior to TT. No strenuous exercise for 24 h prior to TT. On TT day: fasted for 2 h.	Nonnutritive sweetener (Splenda, McNeil Nutritionals, Australia), gelatin capsule.	200 mg (absolute not relative dose; equates to 2.5 mg/kg BW)	Single	No-Doz Awakeners (Key Pharmaceuticals, New South Wales, Australia), gelatin capsule.	50 min prior to exercise (60 min prior to TT).
Stadheim (2013)	No caffeine 48 h prior to TT (with exception for the one high-caffeine consumer who was allowed to consume one-fourth of his normal caffeine amount 48 h to 24 h before testing but refrained from caffeine in the last 24 h). Maintain the same diet for the 48 h before TT. Maintain the same training for the 48 h before TT. On TT day: last meal (a self-selected meal, rich in CHO and protein) was consumed ~1.5 h prior to arriving at the laboratory.	"Vehicle only"	6	Single	Beverage: caffeine (Coffeinum; Oslo Apotekproduksjon, Oslo, Norway) dissolved in a cordial concentrate (Fun Light, 3 mg/mL). Prepared by Ullevål Apotek Produksjon (Oslo, Norway).	75 min prior to TT (45 min prior to warm up).
Stadheim (2015)	No caffeine 48 h prior to TT. Repeat the same diet before all tests (time frame: ns).	"Vehicle only"	4.5	Single	Beverage: caffeine (Coffeinum; Oslo	75 min prior to TT (45 min prior to

	Only light training (and no strength training) allowed in the 48 h before TT. On TT day: ns.				Apotekproduksjon, Oslo, Norway) dissolved in a cordial concentrate (Fun Light, 3 mg/mL). Prepared by test leader.	warm up).
van Nieuwenhoven (2005)	Repeat the same food and fluid on day of each TT. Exercise: ns. On TT day: ns (FSANZ assumes subjects were fed since the TT was at night but it is unclear within what time period prior to TT).	Beverage: CHO (68.8 g/L) and electrolyte drink. Volume: four x 150 mL.	90 mg (absolute not relative dose; equates to mean 1.25 mg/kg BW using mean BW of 72 kg).	Split (> 2 doses)	Beverage: same as C plus caffeine (150 mg/L; equates to 90 mg total).	150 mL 10 min before TT, and 150 mL at 4.5 km, 9 km and 13.5 km during the TT.
Wemple (1997)	No caffeine 4 d prior to TT. Repeat the same diet for 1 d prior to each TT. Exercise: ns. On TT day: fasted overnight, followed by a standardised meal: a 236 mL, 360 kcal liquid meal consisting of 65 % CHO, 17 % fat, and 18 % protein (Gator Pro, Quaker Oats Co., Chicago, IL), 3.5-4 h before the TT.	Beverage: CHO (unknown concentration) and electrolyte drink (Gatorade, Quaker Oats Co., Chicago, IL).	8.7	Split (> 2 doses)	Beverage: same as C plus caffeine (25 mg/dL 1.3.7-trimethylxanthine; Schweizerhall, Inc, Piscataway, NJ). Total caffeine: range: 490-680 mg or 8.7 mg/kg BW.	8 mL/kg BW at t= 0 min, and then 3 mL/kg BW at t=60, 80, 100, 120, 140, 160, 180, 200, and 220 min (where t= 60 min is the start of exercise and t= 250 min is the start of TT). (Total volume: 35 mL/kg BW.)
Womack (2012) AA homozygotes subgroup	No caffeine 24 h prior to TT. Maintain the same training over the course of the study. On TT day: fasted overnight for 12 h.	White flour, capsule.	6	Single	Anhydrous caffeine, capsule.	1 h prior to TT.
Womack (2012) C allele carriers subgroup	Diet: aa. Exercise: aa. On TT day: aa.	aa	6	aa	aa	aa

BW, body weight (kg); ns, not specified; aa, as above; h, hour; d, day; Y, yes; N, no; TT, time trial; CHO, carbohydrate; m, metre; km, kilometre; SS, steady state; HR, heart rate; W_{max} or PO, power output; PPO; peak power output; HIE; high intensity epochs; SE, standard error; t, time; I, intervention condition; C, comparator condition.

[1] Description of trained versus untrained represents the investigators' categorisation. Our categorisation is the same for Astorino (2011 and 2012a) but different for O'Rourke (2008) which we classed as 'uncategorised' due to no VO_{2max} and VO_{2peak} being reported.

[2] Some intervention conditions also included a placebo intake (e.g. non-caffeinated beverage or capsule), which are not listed here.

Table 4: Time trial characteristics

First author (year)	Type, instrument	Familiarisation	Warm up and time trial specifications	Feedback during TT	Washout duration and treatment order effects (if reported)
Acker-Hewitt (2012)	Cycle, computer-simulated TT (Velotron, Racermate, Inc.)	Y	20 min warm up (SS at 60% W_{max}), 20 km TT.	Subjects did not receive encouragement or feedback during this portion of the trial but were provided real-time access to remaining distance, percentage grade, and gearing.	5–14 d
Astorino (2011) and Astorino (2012a) Endurance-trained subgroup[1]	aa	Y	Warm up 50 and 100 W (each 5 minutes), 10 km TT.	Feedback provided on cadence, gearing, and progress on the course via computer screen.	≥48 h No effects of treatment order (P=0.02).
Astorino (2011) and Astorino (2012a) Recreationally active subgroup[1]	aa	aa	aa	aa	≥48 h No effects of treatment order (*P*=0.02).
Astorino (2012b)	aa	Y	aa	aa	48 h–1 week
Astorino (2012c)	aa	Y	8.2 km TT	aa	≥48 h Repeated measures ANOVA with treatment order as a between-subjects factor showed no effect of order on performance (*P*=0.17).
Bell (2002)	Run (while wearing a helmet and backpack weighing 11 kg), treadmill	Y	10 km TT	ns	ns
Bortolotti (2014)	Cycle, computer-simulated TT (Velotron, Racermate, Inc.)	Y	10 min warm up, 20 km TT.	All participants received feedback on the time, power, RPM and distance travelled during the test on a monitor.	≥72 h
Carr (2011)	Row, ergometer (Concept IID, Concept, Vermont, USA)	N (but participants completed a baseline test).	7 min warm up (4 min at 70% of maximal PO, 3-min period of passive rest and 2 × 10 maximal strokes), 2000 m TT.	Subjects performed the TT at the same time as at least 1 other subject to simulate racing conditions.	48 h between first block of three (one block = 1 baseline & 2 experimental trials), and 14-days between block 1 and

					block 2.
Cohen (1996)	Run, road.	N	21 km TT (outdoors in hot and humid conditions. All TTs were performed under high risk heat stress.	Subjects were informed that monetary incentives.	2 weeks
Conway (2003)	Cycle, ergometer (Lode Excalibur Sport, Groningen, The Netherlands)	N	5 min warm up (at a workload of half that calculated to elicit an oxygen uptake of 70% VO_{2max}), 90 min at 68% or 70% VO_{2max} (two different %s are reported in the publication), 30 min TT of a set amount of work (equivalent to 80% VO_{2max} for 30 min).	Subjects were aware of the end point and were verbally encouraged.	7 d Repeated measures ANOVA revealed no main effect of sequence of ingestion on performance time (*P*=0.080).
Cox (2002) Study A	Cycle, ergometer (Lode Instruments, Groningen, The Netherlands)	N	120 min warm up (SS at ~70% VO_{2peak}), TT (7 kJ/kg).	A financial incentive was provided for the fastest TT. The same researcher provided standardised feedback. Subjects only knew the elapsed work as a percentage of the final work; furthermore, subjects were given the results of their TT only after the entire study was completed.	ns
Desbrow (2009)	aa	Y	aa	A financial incentive was provided to produce the fastest TT. The same researcher provided standardised feedback. Subjects were able to view their HR, cadence, and PO for the first 10% of the TT only. After completion of the first 10%, the only information available to subjects was elapsed work as a percentage of the final work; furthermore, subjects were given the results of their TT only after the entire study was completed.	≥7 d
Desbrow (2012)	Cycle, ergometer.	N (the experimental protocol mentions "familiarisation" but it is unclear if this replicated the TT or an ergometer protocol designed to decide individuals' TT parameters).	Standardised warm-up (time: ns), TT of a set amount of work (total work, J = 0.75*PPO*3600).	The same researcher provided standardised feedback to each participant. Participants were able to view their HR, cadence and PO for the first 10% of the time trial only. After completion of the first 10% the only information available to participants was elapsed work as a percentage of the final work.	7 d
Felippe (2018)	Cycle. All trials were performed using the	Y	Standardised warm-up for knee extension muscles (4 x 5 s isometric	Subjects received visual feedback for distance completed but not for exercise time, PO, and pedal frequency.	7 d

	participant's own bike attached to a Computrainer (RacerMate, Seattle, WA)		contractions at 50, 60, 70, and 80% of their maximum voluntary contraction, interspersed by a 30 s rest). Then 45min rest, a repeat of knee extension muscle warm up, 5 min rest, 5 min warm up at 100 W, 5 min rest, 4 km TT.		
Glaister (2015)	Cycle, racing bike (Claud Butler San Remo; Claud Butler, Brigg, United Kingdom) fitted to turbo trainer (Tacx Fortius, Wassenar, The Netherlands).	Y	5 min warm up at 100 W, 5 min rest, 20 km TT.	Verbal encouragement was provided throughout the trial. All measures of elapsed time were removed from the testing environment, and the only data visible to the subjects throughout each TT were the distance completed.	ns
Gonçalves (2017)	Cycle, ergometer (Excalibur; Lode, Groningen, The Netherlands)	Y	5-min warm up at 125 W, TT of a set amount of work (0.85*W_{max}*1800 seconds).	No encouragement was provided during the tests. The only information the participants received during the test was the percentage of work performed relative to the pre-set task, following 25, 50, 75, 90, and 100% completion of the total work done.	≥7 d No order effect was shown (*P*=0.461).
Graham-Paulson (2016)	Cycle (Viking Jetstream 14 road bike), mounted on a Cyclus II ergometer (Avantronic Richter, Leipzig, Germany).	Y	30 min warm up (65% mode-specific V_{peak}), 10 km TT.	No motivation was provided. The only feedback provided was cumulative distance covered.	≥48 h No significant influence of trial order (*P*=0.164).
Guest (2018)	Cycle (Ergomedic 839 E stationary bike)	N (but a 'visit variable' was used to statistically control for potential confounding due to a learning effect).	7 min warm up (light cycling and stretching), vertical jump test, handgrip test, Wingate test, 10 km TT (resistance was set at 65% W_{power}, derived from the VO_{2peak} test).	Subjects were blinded to time, speed, and HR but were able to see distance travelled.	~1 week Potential learning effect was statistically controlled for.
Hodgson (2013)	Cycle, ergometer (Lode Excalibur Sport, Groningen, The Netherlands)	Y	30 min warm up (SS cycling at 50% W_{max} or ~55% VO_{2max}), followed by a TT of a set amount of work (650 ± 37 (SE) kJ at 70% W_{max}).	Subjects did not received verbal or visual feedback regarding performance time or physiological measures throughout the test. Participants received no feedback about their performance until they had completed all trials.	7 d
Hulston & Jeukendrup (2008)	aa	Y	105 min warm up (SS cycling at 62% VO_{2max}), TT of a set amount of work (688 ± 56 (SE) kJ).	Subjects were not given any feedback on their performance until completion of the entire study. Researchers tried to minimise possible distractions.	≥7 d
Hunter (2002)	Cycle, Kingcycle ergometry system (Kingcycle Ltd., High Wycombe, UK), which allows	Y	Isometric testing, 1 h rest, 5 min warm up of "easy cycling", 100 km TT that included bouts of 1- and 4-km	Subjects viewed a diagram of the "course profile", which graphically illustrated where the 1-km and 4-km HIE occurred, before and during each ride. Otherwise subjects received no external clues other than	6–8 d

	cyclists to ride on their own racing bicycles in the laboratory.		HIE.	their elapsed distance and HR. Subjects were not informed of the elapsed time or the times for the HIE until completion of all trials.	
Jacobson (2001)	Cycle, ergometer (Lode Instruments, Groningen, The Netherlands)	N	Standardised 5 min incremental warm up, 120 min SS cycling at 63% peak sustained PO (~70% VO_{2max}), TT at 82.5% peak sustained PO (~85% VO_{2max}) of a set amount of work (7 kJ/kg).	ns	ns
Kovacs (1998)	Cycle, ergometer (Lode Excalibur, Groningen, The Netherlands)	N	20 min warm up, 35 min psychological test (at rest), 5 min warm up, TT of a set amount of work (Joule = 0.75*W_{max}*3600).	Subjects received no information on performance time, work load, pedalling rate, and HR.	7 d
MacIntosh (1995)	Swim, pool.	Y	20 min warm-up, 10 min rest, 1500 m TT in a 25 m pool.	Pacing clocks were shut off and subjects were not permitted to wear wristwatches. Swimmers were assigned lanes, and starts were staggered to minimise the possibility of external pacing cues. No more than 4 swimmers were tested in the 8-land pool at one time.	2–3 d
Miller (2014)	Cycle, ergometer (Lode Excalibur, Groningen, The Netherlands)	Y	80 min training session at 65% VO_{2max}, followed by TT (5 kJ/kg).	ns	2–14 d No order effects were found between treatments for TT performance (*P*=0.776).
O'Rourke (2008) Well-trained subgroup[1]	Run, track	N	Warm up (low to moderate cardiovascular exercise and stretching for ~10-15 min), 5 km TT (around a 400 m athletics track).	Recorded times for all participants were documented and not discussed with the participants until completion of all trials.	ns
O'Rourke (2008) Recreationally active subgroup[1]	aa	Y	aa	aa	aa
Pitchford (2014)	Cycle, ergometer (Lode Excalibur Sport, Groningen, The Netherlands)	Y	TT of a set amount of work (total work (J) = 0.75 × PPO × 2880) in hot conditions, 35°C and 25% relative humidity, via climate chamber).	ns	≥7 d
Potgieter (2018)	Triathlon, field trial.	N	TT. 1.5-km swim, 40-km cycle, and 10-km run, adhering to standard International Triathlon Union guidelines, at Gordons Bay beach, Western Cape, South Africa. Weather	ns	14 d

			posed no risk of heat stress. Entry was restricted to study participants.		
Quinlivan (2015)	Cycle, ergometer (Lode Instruments, Groningen, The Netherlands)	Y	Warm up (self-selected), TT of a set amount of work at 75% PPO.	ns	≥7 d
Roelands (2011)	Cycle, ergometer (Lode Excalibur Sport, Groningen, The Netherlands)	Y	60 min warm up at 55% W_{max}, TT of a set amount of work at ~75% W_{max}.	Percentage of total work completed provided to subjects. No feedback was provided regarding time elapsed, PO, pedal cadence or HR.	7 d
Santos (2013)	Cycle simulator (Tacx Flow T1680, Tacx, Wassenaar, The Netherlands)	Y	5 min warm up at 100 W, 5 min rest, 4000 m TT.	Feedback about the distance covered was provided verbally every 200 m.	7 d Effects of treatment order: none (P>0.05) for ten performance and physiological parameters during the TT. However, time was not reported.
Scott (2015)	Row, ergometer (Concept II, Concept, Vermont, USA)	Y	2 min warm up (self-paced), 1 min rest, 2000 m TT.	Participants were only allowed to see distance left to complete and kcal expended, and were blinded of the time taken to complete the 2000 m until all trials were completed.	3–14 d
Skinner (2010)	aa	Y	20 min warm-up, 2000 m TT.	The performance feedback viewed by participants was PO, average PO, and the distance remaining.	≥7 d
Skinner (2013)	Cycle, ergometer (Wattbike Ltd., Nottingham, England)	Y	15 min warm up, 40 km TT.	Participants knew the distance remaining every 4 km during TT and every 1 km within the last 10% of the workload.	≥5 d
Spence (2013)	Cycle, ergometer (Evolution Pty. Ltd., Adelaide, Australia) attached to custom-designed software (Cylemax, The University of Western Australia).	Y	8 min warm up, 2 min rest, 40 km TT of a set amount of work (1200 kJ).	Subjects were provided feedback indicating the percentage of total work completed (every 5%), but were blinded to all other data output.	7 d
Stadheim (2013)	Cross country double-poling (skiing), ergometer (Thoraxtrainer Elite; Thoraxtrainer, Holbaek, Denmark)	Y	26 min warm up, 5 min rest, 8 km TT.	Subjects could see the remaining distance (m), and were encouraged by a blinded test leader.	6 d
Stadheim (2015)	aa	Y	~25 min warm up, 5 min rest, 8 km TT (during acute hypoxia via hypobaric chamber corresponding to ~2000 m above sea level).	Encouragement was given by a blinded test leader.	7 d
van Nieuwenhoven	Run, field trial	N	18 km TT	ns	3–4 d

(2005)					
Wemple (1997)	Cycle, ergometer	N	180 min warm up (at 60% VO_{2max}), 10 min rest, TT (subjects attempted to complete 500 revolutions on the ergometer as quickly as possible at a pedal resistance that would have elicited 85% VO_{2max} at 75 rpm).	ns	≥4 d
Womack (2012) AA homozygotes subgroup	Cycle, ergometer (Velotron; Racermate, Seattle, WA) on a computer-simulated course.	N	40 km TT	Subjects were blinded to their time, speed, and PO.	ns
Womack (2012) C allele carriers subgroup	aa	aa	aa	aa	aa

BW, body weight (kg); ns, not specified; aa, as above; Y, yes; N, no; TT, time trial; SS, steady state; HR, heart rate; PO, power output; W_{max}, peak power at VO_{2max}; PPO; peak power output; HIE; high intensity epochs; SE, standard error.

[1] Description of trained versus untrained represents the investigators' categorisation. Our categorisation is the same for Astorino (2011 and 2012a) but different for O'Rourke (2008) which we classed as 'uncategorised' due to no VO_{2max} and VO_{2peak} being reported.

Table 5: Outcome (time) data extracted from included studies[1]

First author (year)	Intervention dose (mg/kg BW)	Comparator (placebo)			Intervention (sole or lowest dose)			Intervention (higher dose)			Intervention (highest dose)		
		n	mean	SD	n	mean	SD	n	mean	SD	n	mean	SD
Acker-Hewitt (2012)	6	10	44.2 min	4.5	10	43.6	4.9						
Astorino (2011) and Astorino (2012a) Endurance-trained subgroup[2]	5	8	17.35	0.98	8	Trial 1: 17.07 Trial 2: 17.01 Mean: 17.04	Trial 1: 0.99 Trial 2: 1.0 Mean: 1.00						
Astorino (2011) and Astorino (2012a) Recreationally active subgroup[2]	5	8	18.71	0.68	8	Trial 1: 18.53[3] Trial 2: 18.65 Mean: 18.59	Trial 1: 0.61 Trial 2: 0.80 Mean: 0.71						
Astorino (2012b)	5	9	17.25	0.96	9	Trial 1: 16.98 Trial 2: 16.92 Mean: 16.95	Trial 1: 0.96 Trial 2: 0.97 Mean: 0.97						
Astorino (2012c)	6	10	18.2	1.1	10	17.7	1.0						
Bell (2002)	4	12	46.8[4]	3.2	12	46.0	2.8						
Bortolotti (2014)	6	13	2191 s	157.6 s	13	2181 s	193.9 s						
Carr (2011)	6	8	6:43.8 min:s	23.4 s	8	6:40.8 min:s	22.5 s						
Cohen (1996)	5 or 9	7	88.6[5]	11.4[5]	7	88.3[5]	9.8[5]	7	88.6[5]	101.5[5]			

Conway (2003)	6 (single dose) or 6 split dose (3 before and 3 during the trial)	8	29.3[5]	2.8[5] (SE) 7.92 (SD)	8	24.1[5]	2.8[5] (SE) 7.92 (SD)	8 (split dose)	24.8[5] (split dose)	3.0[5] (SE) (split dose) 8.49 (SD)			
Cox (2002) Study A	6	12	29:18	0:44 (SE) 2:32 (SD)	12	28:18	0:40 (SE) 2:18 (SD)						
Desbrow (2009)	1.5 or 3	9	30 min 25 s	3 min 10 s	9	30 min 42 s	3 min 41 s	9	29 min 51 s	3 min 38 s			
Desbrow (2012)	3 or 6	16	3902 s	340 s	16	3738 s	286 s	16	3791 s	281 s			
Felippe (2018)	5	11	403 s	6 s (SE) 19.90 s (SD)	11	396 s	5 s (SE) 16.58 s (SD)						
Glaister (2015)	5	14	35.37	1.70	14	34.62	1.26						
Gonçalves (2017)	6	40	30.81	2.67	40	29.92	2.18						
Graham-Paulson (2016)	4	11	1016 s	58 s	11	995 s	46 s						
Guest (2018)	2 or 4	101	18.1	0.1 (SEM) 1.00 (SD)	101	17.8[5]	0.1[5] (SEM) 1.00 (SD)	101	17.6	0.3 (SEM) 3.01 (SD)			
Hodgson (2013)[6]	5	8	40.06	0.39 (SEM) 1.10 (SD)	8	38.35	0.48 (SEM) 1.36 (SD)						
Hulston & Jeukendrup (2008)	5.3	10	45.45	1.07 (SEM) 3.38 (SD)	10	43.45	0.86 (SEM) 2.72 (SD)						
Hunter (2002)	9	8	157.7[5]	14.0[5]	8	155.6[5]	15.4[5]						
Jacobson (2001)	6	8	30.37	7.42	8	29.12	5.62						
Kovacs (1998)	2.1, 3.2, or 4.5 (mean)	15	61.5	1.1 (SE) 4.26 (SD)	15	60.4	1.0 (SE) 3.87 (SD)	15	58.9	1.0 (SE) 3.87 (SD)	15	58.9	1.2 (SE) 4.65 (SD)
MacIntosh (1995)	6	11	21:21.8	0:38.2 (SEM) 02:06.7 (SD)	11	20:58.8	0:36.4 (SEM) 02:00.7 (SD)						
Miller (2014)	6 (split dose)	6	20.5	3.5	6 (split dose)	19.7 (split dose)	3.3 (split dose)						
O'Rourke (2008)	5	15	1058 s	68 s	15	1047 s	69 s						

Well-trained subgroup[2]													
O'Rourke (2008) Recreationally active subgroup[2]	5	15	1298 s	84 s	15	1286 s	86 s						
Pitchford (2014)	3	9	4079 s	333 s	9	3806 s	359 s						
Potgieter (2018)	6	26	151.5	18.6	26	149.6	19.8						
Quinlivan (2015)	3	11	3877 s	260 s	11	3757 s	278 s						
Roelands (2011)	6	8	36.6	3.3	8	37.7	5.2						
Santos (2013)	5	8	419 s	13 s	8	409 s	12 s						
Scott (2015)	1.3 (mean)	13	471.4 s	28.5 s	13	466.2 s	26.6 s						
Skinner (2010)	2, 4 or 6	10	403.8	21.0	10	402.4	19.4	10	401.1	19.8	10	402.6	21.2
Skinner (2013)	6	14	3546.2 s	122.8 s	14	3475.7	97.2						
Spence (2013)	2.5 (mean)	10	4497 s	153 s (SEM) 483.83 (SD)	10	4439 s	153 s (SEM) 483.83 (SD)						
Stadheim (2013)	6	10	34:26	1:25 (SEM) 4:28 (SD)	10	33:01	1:24 (SEM) 4:25 (SD)						
Stadheim (2015)	4.5	13	33.25	2.95	13	32.94	2.86						
van Nieuwenhoven (2005)	1.25 (mean)	98	1:18:23	08:47	98	1:18:03	08:42						
Wemple (1997)	8.7 (mean)	6	343 s	19 s (SE) 46.54 s (SD)	6	344 s	24 s (SE) 58.79 s (SD)						
Womack (2012) AA homozygotes subgroup	6	16	76.1	5.8	16	72.4	4.2						
Womack (2012) C allele carriers subgroup	6	19	72.2	4.2	19	70.9	4.3						

BW, body weight.

[1] Where groups' mean and standard deviation were not reported, we calculated them (see Section 1.3.3) and report the raw and calculated data in black and blue text, respectively.

[2] Description of trained versus untrained represents the investigators' categorisation. Our categorisation is the same for Astorino (2011 and 2012a) but different for O'Rourke (2008) which we classed as 'uncategorised' due to no VO_{2max} and VO_{2peak} being reported.

[3] 18.50 in Astorino et al. (2012a).

[4] Later, Bell et al. (2002) says there were 2 placebo trials and reports two values: 46.9 ± 3.3 and 46.7 ± 3.4.

[5] Data were extracted using the online program WebPlotDigitizer Version 4.4. Conway et al. (2003) reports different values in-text for the comparator condition (28.3 ± 3.1), however, as we must extract the intervention conditions' data using WebPlotDigitizer Version 4.4, for consistency we report the comparator condition's data extracted using

WebPlotDigitizer Version 4.4.

[6] Values in abstract are different (larger SE) but we have extracted data from Table 2 and Figure 4 which appear to have smaller SE.

Table 6: Strengths and limitations of our assessment

For efficiency, this assessment was developed from studies included in published systematic reviews, meaning the extent to which primary research was captured is contingent on those reviews' search strategy, eligibility criteria, and accuracy of screening. Furthermore, we did not include research published after August 2017. Given the volume of included research it is unlikely that the unassessed research would change the effect sizes in a meaningful way. Additionally, we did not assess the impact of caffeinated product consumption on nutritional imbalances such as via the contribution to additional sugar intake, displacement of more healthy alternatives, or adverse effect on one's habitual diet.

Table 6: Strengths and limitations of our assessment

Strengths	Limitations
Effect sizes were estimated using meta-analyses. Pooling data increases our ability to detect a real effect if one exists, improves our estimation of the size and direction of effect together with a 95% level of confidence, and decreases the risk of reviewers' bias in interpreting differing results from individual studies.	The assessment question was not established *a priori*.
Outcomes are assessed across a range of intake levels, from 1.25 mg to 9.0 mg caffeine per kilogram body weight.	Studies were not selected and data were not extracted in duplicate.
The pooled sample is large (n=674).	Risk of bias of primary research was not formally assessed.
Most of the studies included in the meta-analyses were randomised, crossover controlled trials with a placebo control.	We identified primary research via a hand search of the reference lists of six key publications. As a result, we did not include research published after August 2017.
	We did not assess the risk of nutritional imbalances and subsequent impact on health outcomes.