



A Systematic Review and Meta-Analysis of Crossover Studies Comparing Physiological, Perceptual and Performance Measures Between Treadmill and Overground Running

Jayne R. Miller¹ · Bas Van Hooren^{2,3} · Chris Bishop^{1,4} · Jonathan D. Buckley¹ · Richard W. Willy⁵ · Joel T. Fuller⁶

Published online: 8 March 2019
© Springer Nature Switzerland AG 2019

Abstract

Background Treadmills are routinely used to assess running performance and training parameters related to physiological or perceived effort. These measurements are presumed to replicate overground running but there has been no systematic review comparing performance, physiology and perceived effort between treadmill and overground running.

Objective The objective of this systematic review was to compare physiological, perceptual and performance measures between treadmill and overground running in healthy adults.

Methods AMED (Allied and Contemporary Medicine), CINAHL (Cumulative Index to Nursing and Allied Health), EMBASE, MEDLINE, SCOPUS, SPORTDiscus and Web of Science databases were searched from inception until May 2018. Included studies used a crossover study design to compare physiological (oxygen uptake [$\dot{V}O_2$], heart rate [HR], blood lactate concentration [La]), perceptual (rating of perceived exertion [RPE] and preferred speed) or running endurance and sprint performance (i.e. time trial duration or sprint speed) outcomes between treadmill (motorised or non-motorised) and overground running. Physiological outcomes were considered across submaximal, near-maximal and maximal running intensity subgroups. Meta-analyses were used to determine mean difference (MD) or standardised MD (SMD) \pm 95% confidence intervals.

Results Thirty-four studies were included. Twelve studies used a 1% grade for the treadmill condition and three used grades $>$ 1%. Similar $\dot{V}O_2$ but lower La occurred during submaximal motorised treadmill running at 0% ($\dot{V}O_2$ MD: -0.55 ± 0.93 mL/kg/min; La MD: -1.26 ± 0.71 mmol/L) and 1% ($\dot{V}O_2$ MD: 0.37 ± 1.12 mL/kg/min; La MD: -0.52 ± 0.50 mmol/L) grade than during overground running. HR and RPE during motorised treadmill running were higher at faster submaximal speeds and lower at slower submaximal speeds than during overground running. $\dot{V}O_2$ (MD: -1.25 ± 2.09 mL/kg/min) and La (MD: -0.54 ± 0.63 mmol/L) tended to be lower, but HR (MD: 0 ± 1 bpm), and RPE (MD: -0.4 ± 2.0 units [6–20 scale]) were similar during near-maximal motorised treadmill running to during overground running. Maximal motorised treadmill running caused similar $\dot{V}O_2$ (MD: 0.78 ± 1.55 mL/kg/min) and HR (MD: -1 ± 2 bpm) to overground running. Endurance performance was poorer (SMD: -0.50 ± 0.36) on a motorised treadmill than overground but sprint performance varied considerably and was not significantly different (MD: -1.4 ± 5.8 km/h).

Conclusions Some, but not all, variables differ between treadmill and overground running, and may be dependent on the running speed at which they are assessed.

Protocol registration CRD42017074640 (PROSPERO International Prospective Register of Systematic Reviews).

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s40279-019-01087-9>) contains supplementary material, which is available to authorized users.

✉ Joel T. Fuller
joel.fuller@mq.edu.au

Extended author information available on the last page of the article

1 Introduction

Motorised treadmills are widely used by clinicians, researchers and athletes to approximate overground running for diagnosis and rehabilitation of injuries, training, athletic performance testing and research purposes [1–3]. Non-motorised treadmills have also been used for these purposes, particularly when instantaneous changes in running

Key Points

During submaximal motorised treadmill running, oxygen uptake ($\dot{V}O_2$), heart rate and perceived effort were similar, but blood lactate concentration (La) was lower and preferred running speeds were slower than during overground running.

At near maximal running speeds, $\dot{V}O_2$ and La were lower during motorised treadmill running than during overground running but there was no difference in $\dot{V}O_2$ at maximal running speed.

Endurance performance was poorer on a motorised treadmill than overground but sprint performance varied considerably.

speed are required [4], and anecdotal reports suggest they are becoming a popular training device for runners [5]. The main advantage of treadmills, compared with overground running, is that they allow various physiological measures to be captured within a limited and controlled space, greater control over performance variables such as running velocity and surface gradient, and greater control over environmental variables such as ambient temperature, wind speed and relative humidity when compared with outdoor running [6].

Despite the widespread use of treadmills, there remain concerns about whether treadmill running is a valid surrogate of the physiological demands required during overground running. Pugh [7] was the first to investigate this issue and demonstrated a higher rate of oxygen uptake ($\dot{V}O_2$) in one group of runners on an outdoor track compared with a different group of runners on a motorised treadmill. Maksud and colleagues [8] were the first to use a crossover study design and found that minute ventilation was similar, but $\dot{V}O_2$ was generally higher, during overground compared with motorised treadmill running. Both authors suggested that the lack of air resistance during treadmill running was the reason for the $\dot{V}O_2$ differences [7, 8]. However, subsequent studies have reported conflicting findings, with some [9, 10] finding no significant difference in $\dot{V}O_2$ between the two conditions and others [11, 12] reporting higher $\dot{V}O_2$ during motorised treadmill running than during overground running at the same speed. Notably, Jones and Doust [2] found that $\dot{V}O_2$ could be matched between overground and motorised treadmill conditions by adjusting the treadmill grade to 1%. As a result, many researchers adjust the treadmill grade to 1% [12–16], despite not all studies supporting this practice [5, 12]. Using a treadmill grade such as 1% might be particularly important at faster running speeds, which tend to demonstrate greater air resistance during overground running

and hence larger $\dot{V}O_2$ differences between motorised treadmill and overground running than at slower speeds [8, 17].

While $\dot{V}O_2$ is the most commonly compared variable between treadmill and overground running, a range of other variables have also been investigated. Physiological variables such as heart rate (HR) and blood lactate concentration (La) have been investigated, with initial studies suggesting mixed results for blood lactate [18] and minimal differences for HR [8]. Studies investigating running performance have produced conflicting results for maximal 100 m sprint speed [6, 19] and poorer endurance performance (i.e. slower 5 km and 1-h time trials) on a motorised treadmill compared with overground running [16, 20]. Consistent with findings of poorer endurance performance, some studies have reported that motorised treadmill running is perceived as requiring greater effort than overground running [14] and runners select slower running speeds on a motorised treadmill than overground when instructed to run at a fixed level of perceived effort [21].

Despite a considerable amount of original research comparing physiological, perceptual and performance measures between treadmill and overground running, there has been no systematic review and synthesis of the current literature on this topic. Such a review is important, considering the widespread use of treadmills within research and clinical settings and the apparent inconsistent findings of existing literature. Additionally, the ability to more precisely apply laboratory-based findings to overground conditions could be beneficial to researchers, clinicians and athletes in the context of running performance enhancement and rehabilitation. A systematic review would also highlight best practices in data collection procedures so that clinicians can maximise generalisability between treadmill and overground testing. As a result, the aim of this review was to compare physiological, perceptual and performance measures between treadmill and overground running conditions by reviewing crossover studies that have compared these two conditions in healthy adult participants.

2 Methods

2.1 Registration of Systematic Review Protocol

A systematic review of the literature was performed according to the *Cochrane Handbook for Systematic Reviews of Interventions* (version 5.1.0; <http://handbook-5-1.cochrane.org/>) and following the checklist for the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) 2015. The original protocol was prospectively registered with the International Prospective Register of Systematic Reviews (PROSPERO) in October 2017 and involved a

larger systematic review project comparing overground and treadmill running across a broader range of outcome measures (Registration number CRD42017074640). A revised version of the review protocol was registered with Open Science Framework in December 2018 [51] and published with PROSPERO in January 2019 after the scope of the review was narrowed to include only physiological, perceptual and performance outcome measures.

2.2 Information Sources

The following bibliographic databases were searched from inception to May 2018: AMED (Allied and Contemporary Medicine), CINAHL (Cumulative Index to Nursing and Allied Health), EMBASE, MEDLINE, SCOPUS, SPORT-Discus and Web of Science. Database searches were supplemented by forward citation tracking and manual reference list searches of eligible studies. Search alerts were used to monitor for new search results after the date of the initial search (October 2017) until the completion of the review (May 2018). Any relevant articles identified by these alerts were assessed for eligibility by two reviewers (JRM and JTF).

2.3 Eligibility Criteria

Eligibility criteria for the search strategy are shown in Table 1. $\dot{V}O_2$, HR, blood lactate, rating of perceived exertion (RPE) and preferred running speed outcomes were considered submaximal ($< 80\%$ maximal $\dot{V}O_2$ [$\dot{V}O_{2max}$] or La < 4 mmol/L), near-maximal ($\geq 80\%$ $\dot{V}O_{2max}$ or La ≥ 4 mmol/L) or maximal (100% $\dot{V}O_{2max}$) intensity. For studies that did not report percentage $\dot{V}O_{2max}$ ($\% \dot{V}O_{2max}$), the estimated $\% \dot{V}O_{2max}$ was calculated using the reported group mean $\dot{V}O_2$ and $\dot{V}O_{2max}$, information provided by corresponding authors of the included studies or qualitative comparison with other studies involving similar speeds and participants of similar running ability. For studies that compared outcomes between treadmill and overground running in shod and barefoot conditions, we included only data from the shod condition.

2.4 Search Strategy

The following search strategy was adapted for each database and applied to the title, abstract and keyword search fields:

run* and treadmill and [over-ground or over ground or overground or track or field or outdoor] and [rpe or exertion or effort or perce* or psychophysic* or performance* or time* or distance* or speed* or accelerat* or endurance or economy or efficiency or energy or oxygen or vo2 or aerobic or cardiovascular or heart or hr or lactic or lactate or biomechanic* or motion or

Table 1 Eligibility criteria

Criterion	Description
Type of participant	Healthy adults (aged ≥ 18 years), males and females
Type of condition	Motorised and non-motorised treadmill
Type of comparison	Overground (indoors and outdoors)
Type of outcomes	Physiological measures: oxygen uptake, heart rate and blood lactate concentration Perceptual measures: rating of perceived exertion; preferred running speed Performance measures: time trial; maximal sprint speed, maximal aerobic speed
Type of study	Crossover studies
Publication status	Peer-reviewed journal
Publication language	English

camera* or kinematic* or kinetic* or force* or analysis* or 3d or three-dimension* or three dimension* or grf or work or torque* or moment* or impulse or emg or electromyography or muscle* or stride* or step* or pelvis or hip* or knee* or ankle* or thigh* or shank* or leg* or foot*]

Free-text terms were chosen based on pilot searches that achieved an appropriate balance between search sensitivity and precision without the use of controlled vocabulary (e.g. MeSH [Medical Subject Headings]). Searches were limited to English-language articles only, but no limits were placed on publication date. Biomechanical search terms were used because the search was part of a larger systematic review project comparing overground and treadmill running across a range of outcome measures.

2.5 Study Selection

Study eligibility assessments were completed independently by two reviewers (JRM and JTF). Records were downloaded into Endnote (version X8.2, Clarivate Analytics, Philadelphia, PA, USA) and duplicates were removed before being screened by title and abstract. The full texts of the remaining records were then retrieved and assessed for eligibility. Any discrepancies throughout the study selection process were settled through consultation with all reviewers until a consensus decision was reached.

2.6 Data Collection Process

Data extraction was performed independently by three reviewers (JRM, JTF and BvH) using a standardised form. The data were then merged by two authors (JRM and JTF) and any discrepancies were settled through consultation with all authors. The data extracted from each study included

(1) publication details; (2) participant characteristics; (3) study design; (4) experimental conditions; (5) randomisation method; (6) outcome measures; (7) attrition; (8) statistical analysis; (9) mechanism for controlling environmental conditions; and (10) running intensity. If insufficient information was reported, the corresponding authors of the included studies were contacted by e-mail to obtain missing information.

2.7 Risk of Bias Assessment

This review used a modified version of the Cochrane Collaboration tool for risk of bias assessment [22]. Modifications included removal of the performance bias criterion and addition of statistical bias, environmental bias and intensity bias criteria. The performance bias criterion was removed because it was considered impossible to successfully blind participants and personnel to running conditions. Statistical bias related to whether the included studies used appropriate statistical analysis for paired data [23]. Environmental bias was a qualitative assessment of how effectively temperature, humidity, wind and surface gradient were matched between overground and treadmill running conditions. Studies performed outdoors were considered to have a high risk of environmental bias. Intensity bias was a qualitative assessment of how effectively running speed, distance and duration were matched between overground and treadmill running conditions. Risk of bias assessment was performed independently by three reviewers (JRM, JTF and RWW) and any discrepancies were settled through consultation with all authors.

2.8 Statistical Considerations

Random-effects meta-analyses were performed using the Metafor statistical package in R software (version 3.4.3, R Foundation for Statistical Computing). A separate random-effects meta-analysis was performed for each individual review outcome. Subgroup analyses were performed based on treadmill grade and whether the treadmill was motorised or not, and meta-regression was performed using running speed as a continuous predictor variable. Individual study effects were weighted using the inverse variance method. If studies reported a mean difference (MD) and exact p -value without reporting the variance of the MD, we calculated the variance based on the equivalent T -statistic. Where sufficient information was available, the correlation between treadmill and overground running measurements was also estimated. The lowest correlation coefficient for each review outcome measure was then used to estimate the variance of the MD for studies that did not report sufficient information. A correlation

coefficient of 0.50 was used when no correlation estimates were available for a specific review outcome. This process has been described in detail previously [24] and ensured that the maximum number of studies was included in each meta-analysis.

Multiple study effect sizes and variances were included for studies investigating multiple treadmill grades. The study sample was divided amongst the effects to avoid double counting of individuals from the same study. For studies that compared outcomes between treadmill and overground running at multiple running speeds and using multiple running surfaces or treadmill models, we included a single effect size and variance to avoid double counting individuals from those studies in the meta-analysis. The included effect size and variance represented the mean value across the different speeds, surfaces or treadmill models that were compared within the study.

Raw MDs were used for meta-analyses when the included studies reported results for review outcomes using the same measurement scale. Standardised MDs (SMDs) were calculated for review outcomes when the included studies reported results for review outcome using difference measurement scales (i.e. 6–20 vs. 0–10 perceived exertion scales or 5 km vs. 10 km vs. 1-h time trials). Raw MDs were converted to SMDs by dividing by the pooled between-subject standard deviation using the Cohen's d_{average} method with Hedges' correction [25]. If the between-subject standard deviation was not reported, the MD was converted to a SMD by dividing by the pooled between-subject standard deviation from all other included studies that reported the same outcome measure. SMDs were considered trivial (< 0.20), small (0.20–0.59), moderate (0.60–1.19), large (1.20–1.99) and very large (≥ 2.00) [26].

The Cochran's Q and I^2 statistics were used to investigate the presence of statistical heterogeneity within each meta-analysis. Statistical heterogeneity was considered low ($I^2 < 25\%$), moderate ($I^2 25\text{--}49\%$) and high ($I^2 > 50\%$) [27]. Publication bias was not assessed because there was only a small number of studies included in each meta-analysis and there was no reason to expect that studies finding no difference between treadmill and overground would be less likely to be published than studies reporting a statistically significant difference. The overall quality of the evidence synthesis was rated high, moderate, low or very low using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system [28]. The quality rating was downgraded one level from high for each of the following limitations: total sample size < 100 participants (imprecision); high statistical heterogeneity (inconsistency); and more than 50% of studies in meta-analysis had more than one risk of bias item assessed to be high risk.

3 Results

3.1 Search Results

Database searching identified 1138 records after duplicates were removed. An additional eight records were identified through forward citation searching and manual checking of the reference lists of included articles. A summary of the search results and reasons for exclusion are shown in Fig. 1. Thirty-four studies were included in this review and are summarised in Table 2. Non-motorised treadmill running was considered in only three of the 34 included studies [5, 14, 29]. This small number of studies was considered insufficient for undertaking a systematic review. Results of these studies are summarised in Appendix S1 of the Electronic Supplementary Material (ESM).

3.2 Participants

This review contained a total of 468 total participants (350 males, 102 females, 16 not specified). Of the 34 studies included in the review, 23 studies [1, 2, 6, 8–10, 14, 16–19, 21, 29, 30, 32, 34, 35, 37, 38, 40–43] included only male participants, two studies [12, 36] did not specify participant sex, and nine studies [5, 11, 13, 15, 20, 31, 33, 39, 44] included a mix of male and female participants. Twenty-one studies recruited only runners and typically described the runners in their studies as ‘trained’, ‘distance’ or ‘recreational’ runners. The remaining 13 studies included healthy individuals, non-runners or athletes from other sporting

backgrounds. Only one study [29] reported recruiting participants with previous treadmill running experience. Seven studies [13, 15–17, 20, 39, 43] reported the minimum or mean weekly running distance of participants. Hanson and colleagues [13] recruited participants with the lowest weekly running distance (range: 16–40 km per week). Aubry and colleagues [17] reported the highest weekly running distance (mean: 120 ± 32 km per week). Ten studies [2, 5, 9, 10, 13, 17, 30, 38, 41, 44] reported the $\dot{V}O_{2\max}$ of participants at the time of enrolment. Hanson and colleagues [13] recruited participants with the lowest $\dot{V}O_{2\max}$ values (mean [males]: 48.0 ± 3.6 mL/kg/min; mean [females]: 44.4 ± 9.2 mL/kg/min). White et al. [44] recruited participants with the highest $\dot{V}O_{2\max}$ values (mean [males]: 72.7 ± 1.1 mL/kg/min; mean [females]: 64.0 ± 3.3 mL/kg/min).

3.3 Conditions

Of the 34 studies, 12 studies used a 1% grade for the treadmill condition [2, 5, 12–16, 18, 31, 38, 39, 42] and three used grades $> 1\%$ [2, 30, 41]. The remaining studies either stated that no treadmill grade was set or did not mention treadmill grade. Nine studies [5, 8, 11, 13, 14, 33, 38–40] included overground running trials indoors on synthetic, wooden or unspecified surfaces (Table 2). For the 26 studies that included overground trials outdoors, 14 studies [1, 2, 6, 16, 18, 19, 29–32, 34, 35, 39, 41] specified the running surface and the remaining 12 studies [9, 10, 12, 15, 17, 20, 21, 36, 37, 42–44] were described as ‘track’, ‘athletic track’ or ‘outdoors’ (Table 2).

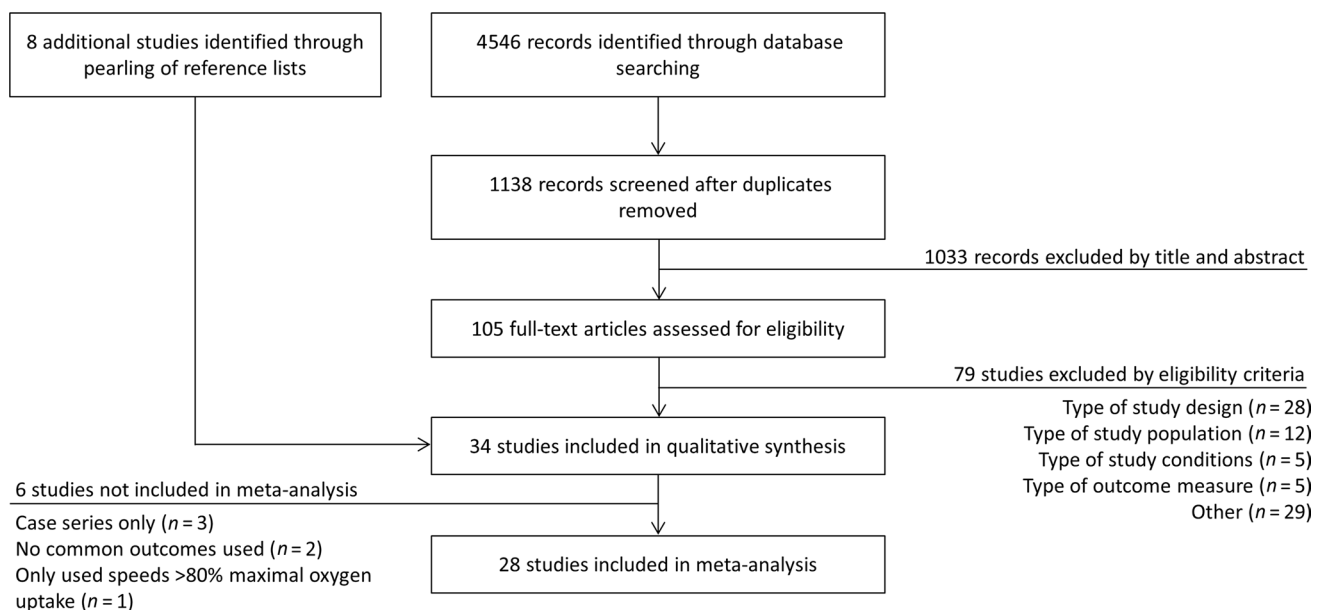


Fig. 1 Literature search flow chart. *n* number of studies

Table 2 Summary of included studies

Study	Year	<i>n</i> (M/F)	Runners	Treadmill (grade)	Overground (surface)	Workload	MT vs. overground
Aubry et al. [17]	2018	24/0	Yes	MT (0%)	Outdoor (track)	2–4 min; 11, 12, 14, 16, 18 km/h	$\dot{V}O_2$ (↓)
Bassett et al. [30]	1985	7/0	No	MT (0%, 5.7%)	Outdoor (track 0%; road 5.7%)	950 m; 8.4–16.8 km/h	$\dot{V}O_2$ (↔)
Bidder et al. [31]	2017	7/13	No	MT (0%, 1%)	Outdoor (tarmac; grass)	4 min; 8 km/h	$\dot{V}O_2$ (↓)
Bonen et al. [32]	1974	2/0	No	MT (0%)	Outdoor (cement; sidewalk; tartan; cinder)	10 min; 10 km/h	$\dot{V}O_2$ (↔)
Bowtell et al. [33]	2009	6/5	No	MT (0%)	Indoor (hard-floor)	10 s sprint	Velocity (↑)
Brookes et al. [34]	1971	3/0	Yes	MT (0%)	Outdoor (road)	9.6, 12.9, 16.1, 19.3 km/h	HR, $\dot{V}O_2$ (↔)
Cappa et al. [1]	2014	14/0	No	MT (0%)	Outdoor (grass)	Maximal GET	MAS (↑)
Ceci and Hassmén [21]	1991	11/0	Yes	MT (0%)	Outdoor (track)	3 min; RPE 11, 13, 15	HR, La, velocity (↓)
Chu et al. [35]	2010	26/0	No	MT (0%)	Outdoor (rubber)	20 min; self-selected speed	HR (↓)
Dal Monte et al. [36]	1974	3 ^a	No	MT (0%)	Outdoor (track)	15, 18, 20 km/h	$\dot{V}O_2$ (↔)
Edwards et al. [5]	2017	7/7	Yes	MT (1%), NMT (unclear)	Indoor (wooden)	6 min; 9, 10.5, 12, 13.5, 15, 16.5 km/h	HR, RPE $\dot{V}O_2$ (↔)
Frishberg [19]	1983	5/0	Yes	MT (0%)	Outdoor (artificial)	91 m sprint	Velocity (↔)
Hanson et al. [13]	2011	5/5	Yes	MT (1%)	Indoor (track)	6 min; 70% $\dot{V}O_{2max}$	HR, $\dot{V}O_2$, RPE(↔)
Harte and Eifert [37]	1995	10/0	Yes	MT (0%)	Outdoor (unclear)	30 min; 15–18 km/h	RPE (↑)
Heck et al. [18]	1985	6/0	No	MT (1%)	Outdoor (rekortan)	3 min; 9.4, 10.8, 12.2, 13.7, 15.1 km/h	La (condition × speed)
Heesch and Slivka [38]	2015	10/0	Yes	MT (1%)	Indoor (rubber)	10 km time trial	Time (↓), HR, RPE (↔)
Jones and Doust [2]	1996	9/0	Yes	MT (0%, 1%, 2%, 3%)	Outdoor (road)	6 min; 10.5, 12.0, 13.5, 15.0, 16.5, 18 km/h	HR, $\dot{V}O_2$ (condition × grade)
LaCaille et al. [39]	2004	22/38	Yes	MT (1%)	Indoor (track); outdoor (road)	5 km time trial	RPE, time (↑)
Maksud et al. [8]	1971	15/0	No	MT (0%)	Indoor (track)	2 min; 11.2, 16.1, 19.3 km/h	HR, $\dot{V}O_2$ (↔)
McMiken and Daniels [9]	1976	8/0	Yes	MT (0%)	Outdoor (track)	6 min; 10.8, 12.6, 15.5 km/h	$\dot{V}O_2$ (↔)
McMurray et al. [10]	1988	8/0	Yes	MT (0%)	Outdoor (track)	16 km; 70% $\dot{V}O_{2max}$	HR, RPE, $\dot{V}O_2$ (↔), La (↓)
Meyer et al. [40]	2003	18/0	No	MT (0.5%)	Indoor (tartan)	Maximal GET	HR _{max} (↓), $\dot{V}O_{2max}$ (↔)
Mooses et al. [12]	2015	13 ^a	Yes	MT (1%)	Outdoor (track)	3 min; 16 km/h (> 85% $\dot{V}O_{2max}$); maximal GET	HR (↔), $\dot{V}O_2$ (↑), $\dot{V}O_{2max}$ (↔)
Morin and Sève [6]	2011	11/0	No	MT (0%)	Outdoor (tartan)	100 m sprint	Time (↑), velocity (↓)
Olivier and Scott. [41]	1993	11/0	Yes	MT (3.8–7.5%)	Outdoor (road 3.8–7.5%)	20 min; 70% $\dot{V}O_{2max}$	HR, $\dot{V}O_2$ (↔), RPE (↑)
Panasci et al. [42]	2017	15/0	No	MT (1%)	Outdoor (track)	15 × 30/30 s HIIT; MAS	HR (↔), RPE, $\dot{V}O_2$ (↓)
Peserico and Machado [16]	2014	18/0	Yes	MT (1%)	Outdoor (woodland trail)	1 h time trial	HR, velocity (↓), HR _{max} (↑)
Ramsbottom et al. [20]	1992	6/6	Yes	MT (0%)	Outdoor (track)	5 km time trial	Time (↑)
Ramsbottom et al. [43]	1992	6/0	Yes	MT (0%)	Outdoor (track)	5 km time trial	Time (↑)

Table 2 (continued)

Study	Year	<i>n</i> (M/F)	Runners	Treadmill (grade)	Overground (surface)	Workload	MT vs. overground
Stevens et al. [29]	2015	10/0	Yes	NMT (curve)	Outdoor (tartan)	5 km time trial	N/A (NMT results only)
Turner and Stevinson [15]	2017	14/8	Yes	MT (1%)	Outdoor (track)	3 km; “steady pace”	RPE (↔)
Wee et al. [14]	2016	10/0	Yes	MT, NMT (1%)	Indoor (mondo track)	1000 m; 12, 14, 16 km/h, maximum	HR, La, $\dot{V}O_2$ (↔), RPE (↑)
White et al. [44]	1998	12/6	Yes	MT (0%)	Outdoor (track)	60 min; 75% $\dot{V}O_{2max}$	HR (↑), La (condition × sex)
Yngve et al. [11]	2003	14/14	No	MT (0%)	Indoor (gym)	5 min; “comfortable pace”	HR (↔), $\dot{V}O_2$ (↑)

F female, *GET* graded exercise test, *HIIT* high-intensity interval training, *HR* heart rate, *HR_{max}* maximum heart rate, *La* blood lactate concentration, *M* male, *MAS* maximal aerobic speed, *MT* motorised treadmill, *n* sample size, *N/A* not applicable, *NMT* non-motorised treadmill, *RPE* rating of perceived exertion, $\dot{V}O_2$ oxygen uptake, $\dot{V}O_{2max}$ maximal oxygen uptake, ↔ indicates no statistically significant difference between treadmill and overground running, ↑ indicates statistically significant increase during treadmill compared to overground running, ↓ indicates statistically significant decrease during treadmill compared with overground running

^aParticipant sex unclear

3.4 Running Workload

Fourteen studies involved participants running at a fixed running speed [2, 5, 8, 9, 12, 14, 17, 18, 30–32, 34, 36, 37], four studies involved participants running at a fixed self-selected pace [11, 15, 35] or RPE [21] and four studies involved participants running at a fixed percentage of $\dot{V}O_{2max}$ (70–75% $\dot{V}O_2$) [10, 13, 41, 44]. Six studies involved set distance or duration time trials [16, 20, 29, 30, 38, 43], three studies involved maximal effort sprint protocols [6, 19, 33] and two studies involved maximal effort graded exercise tests [1, 40]. One study involved repeated high-intensity running bouts [42].

3.5 Risk of Bias

All eligible studies used a crossover design. Sixteen studies [1, 10–12, 14, 16, 17, 19, 20, 30, 32–34, 36, 41, 43] were at high risk of an order effect, reporting a fixed starting condition or not reporting any randomisation. Only three studies [13, 15, 44] were classified as low risk, providing sufficient information regarding method of randomisation. All other studies were classified as unclear risk because they did not provide sufficient information regarding method of randomisation. No studies provided information regarding allocation concealment. No studies referred to publicly available study protocols, so it was unclear whether results were affected by selective reporting. Only three studies [8, 14, 38] conducted all trials in the same testing area indoors and were at low risk of environmental bias. Twenty-six studies conducted their overground trials outdoors and were classified as having high risk of bias due to difficulties controlling wind speed, relative humidity and ambient temperature. The remaining five studies [5, 11, 13, 33, 40] either

did not provide sufficient information to determine whether the overground condition took place indoors or outdoors or did not report sufficient information regarding control of indoor environmental conditions. Eight studies [2, 10, 14, 17, 31, 39, 40, 42] did not provide sufficient information on the number of participants assessed and included in analysis and were at an unclear risk of attrition bias. Five studies [8, 15, 20, 30, 33] were at high risk of attrition bias, reporting incomplete data for all subjects or reporting study dropouts without sufficient reason. Five studies [15, 18, 19, 32, 36] did not provide sufficient information regarding how they controlled the speed for the overground trials and were classified as at unclear risk of intensity bias. Eight studies [17, 21, 30, 34, 37, 39, 41, 42] were at high risk of intensity bias because they did not appropriately control running intensity between conditions or due to discrepancies in the protocol for each condition. There was a low risk of statistical bias in 25 studies and unclear risk in six studies [5, 6, 9, 11, 17, 42]. Risk of statistical bias was high in one study [33] because the authors reported using a Mann–Whitney *U* test instead of a Wilcoxon signed-rank test to compare paired data, and in two other studies [34, 36] because no inferential statistics were attempted due to sample size.

3.6 Oxygen Uptake

3.6.1 Submaximal Running

Pooled results indicated that submaximal motorised treadmill running at 0% grade reduced $\dot{V}O_2$ by 0.55 mL/kg/min and 1% grade increased $\dot{V}O_2$ by 0.37 mL/kg/min (Fig. 2) compared with overground running. These pooled differences were affected by high heterogeneity ($I^2 > 75\%$) and were not significant at 0% ($p = 0.251$;

$n = 116$; number of independent studies [k] = 9) or 1% relative grade ($p = 0.695$; $n = 48$; $k = 5$). The GRADE quality of evidence was low for the 0% and 1% subgroups and moderate for the overall motorised treadmill pooled results (Table 3). Motorised treadmill running speed was not a significant moderator of the MD in submaximal $\dot{V}O_2$ ($\beta = -0.03$ mL/kg/min; 95% confidence interval [CI] -0.52 to 0.47 ; $p = 0.914$; Fig. 3).

One study investigating the effect of treadmill grades ranging from 0% to 3% in nine runners indicated that 1% grade resulted in the most closely matched $\dot{V}O_2$ between submaximal motorised treadmill and overground running [2]. Three case series with only two to three runners or non-runners concluded that $\dot{V}O_2$ was similar during submaximal overground and motorised treadmill running at 0% [32, 34, 36].

3.6.2 Maximal Running

Pooled results indicated that near-maximal ($\geq 80\%$ $\dot{V}O_{2max}$) motorised treadmill running at 0% grade significantly reduced $\dot{V}O_2$ by 3.12 mL/kg/min ($p = 0.011$; $n = 44$; $k = 3$) compared with overground running but there was no significant difference at 1% grade ($p = 0.898$; $n = 45$; $k = 5$) (Fig. 4). These pooled differences were affected by high heterogeneity ($I^2 > 87\%$). The GRADE quality of evidence was low for the 0% and 1% subgroups and the overall motorised treadmill pooled results (Table 3). Pooled results indicated that maximal motorised treadmill running caused non-significant 0.78 mL/kg/min increases in $\dot{V}O_2$ compared with overground running ($p = 0.323$; $n = 29$; $k = 3$) with moderate heterogeneity ($I^2 = 27\%$; Fig. 4) and moderate GRADE quality of evidence (Table 3).

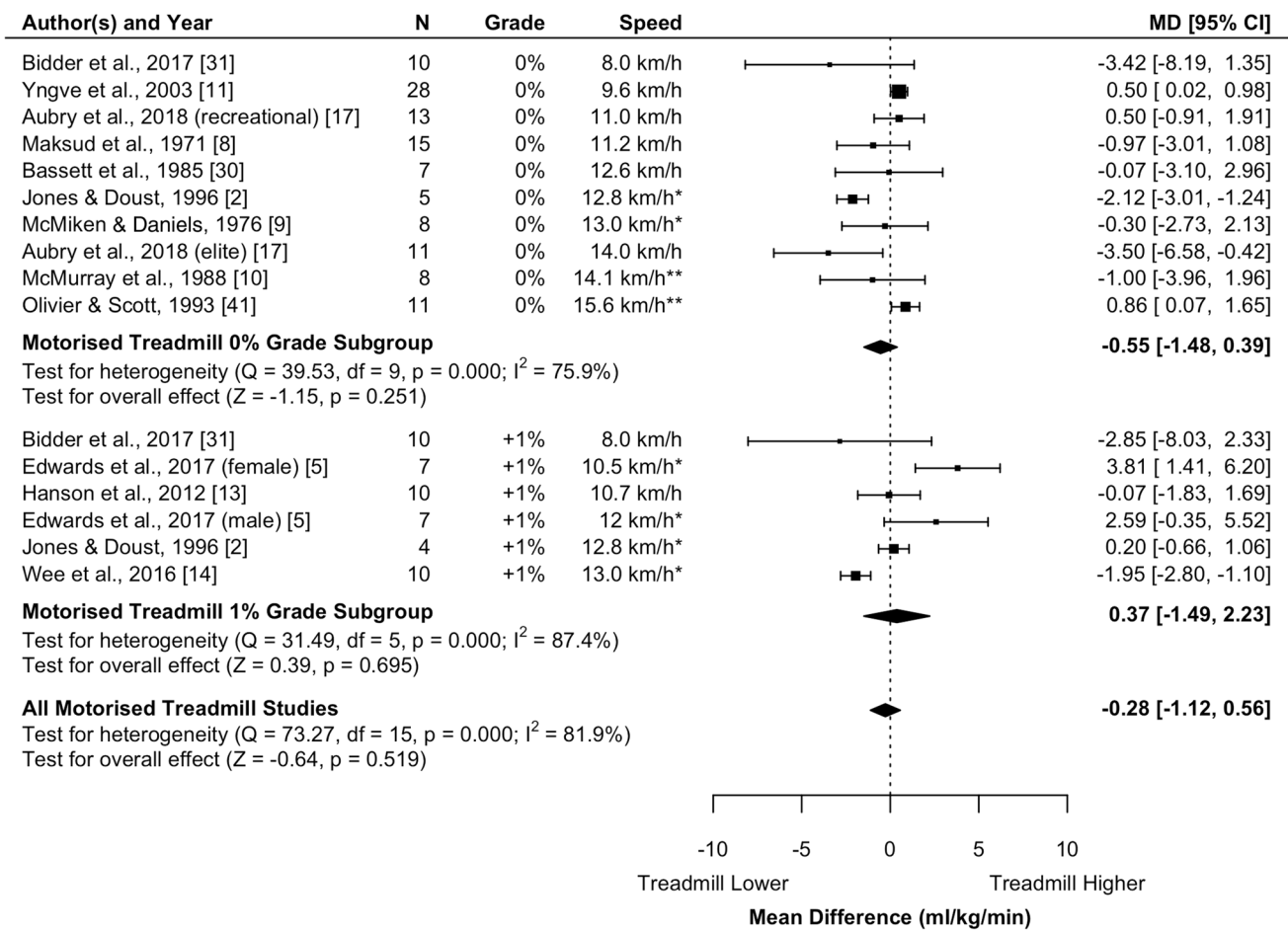


Fig. 2 Random-effects meta-analysis of submaximal oxygen uptake during treadmill compared to overground running. Subgroup analysis based on relative treadmill grade. *Mean submaximal running speed used in study. **Running speed estimated from the Jones and Doust

[2] study that involved runners of similar running ability. *CI* confidence interval, *df* degrees of freedom, *MD* mean difference, *N* sample size

Table 3 Summary of motorised treadmill findings and quality of evidence synthesis

Outcome	Summary of findings			Quality of evidence synthesis (GRADE)			
	<i>k</i>	<i>n</i>	Effect (95% CI)	Imprecision	Inconsistency	Risk of bias	Overall quality
Submaximal $\dot{V}O_2$ (mL/kg/min)							
0% grade	9	116	-0.55 (-1.48 to 0.39)	None	-1	-1 ^a	Low
1% grade	5	48	0.37 (-1.49 to 2.23)	-1	-1	None	Low
All studies	12	164	-0.28 (-1.12 to 0.56)	None	-1	None	Moderate
Near-maximal $\dot{V}O_2$ (mL/kg/min)							
0% grade	3	44	-3.12 (-5.12 to -0.72)	-1	-1	None	Low
1% grade	5	45	0.18 (-2.64 to 3.01)	-1	-1	None	Low
All studies	7	89	-1.25 (-3.34 to 0.85)	-1	-1	None	Low
Maximal $\dot{V}O_2$ (mL/kg/min)							
All studies	3	29	0.78 (-0.77 to 2.33)	-1	None	None	Moderate
Submaximal HR (bpm)							
0% grade	8	122	-3 (-12 to 6)	None	-1	None	Moderate
1% grade	4	38	3 (0 to 5)	-1	None	None	Moderate
All studies	12	160	-1 (-6 to 5)	None	-1	None	Moderate
Near-maximal HR (bpm)							
0% grade	2	20	-3 (-6 to -1)	-1	None	None	Moderate
1% grade	5	45	0 (-1 to 2)	-1	None	None	Moderate
All studies	6	65	0 (-2 to 1)	-1	-1	None	Low
Maximal HR (bpm)							
All studies	2	23	-1 (-3 to 1)	-1	None	None	Moderate
Submaximal blood lactate (mmol/L)							
0% grade	3	37	-1.26 (-1.97 to -0.56)	-1	None	-1 ^a	Low
1% grade	2	16	-0.52 (-1.02 to -0.02)	-1	None	None	Moderate
All studies	5	53	-0.95 (-1.50 to -0.40)	-1	-1	None	Low
Near-maximal blood lactate (mmol/L)							
All studies	2	20	-0.54 (-1.17 to 0.09)	-1	None	None	Moderate
Submaximal RPE (SMD)							
0% grade	4	55	0.19 (-0.81 to 1.18)	-1	-1	-1 ^a	Very low
1% grade	4	56	0.10 (-0.40 to 0.60)	-1	-1	None	Low
All studies	8	111	0.16 (-0.31 to 0.63)	None	-1	None	Moderate
Near-maximal RPE (MD)							
All studies	3	39	-0.44 (-2.48 to 1.60)	-1	-1	None	Low
Sprint performance (km/h)							
0% grade	2	16	-3.5 (-10.3 to 3.4)	-1	-1	-1 ^a	Very low
All studies	3	27	-1.4 (-7.1 to 4.4)	-1	-1	-1 ^a	Very low
Endurance performance (SMD)							
0% grade	2	18	-0.55 (-0.82 to -0.29)	-1	None	-1 ^a	Low
1% grade	3	88	-0.39 (-1.02 to 0.25)	-1	-1	-1 ^a	Very low
All studies	5	106	-0.50 (-0.86 to -0.14)	None	-1	-1 ^a	Low

CI confidence interval, *GRADE* Grading of Recommendations Assessment, Development and Evaluation, *HR* heart rate, *k* number of independent studies, *MD* mean difference, *n* number of participants, *RPE* rating of perceived exertion, *SMD* standardised mean difference, $\dot{V}O_2$ oxygen uptake

^a>50% of studies with >1 risk of bias assessment item rated as high-risk

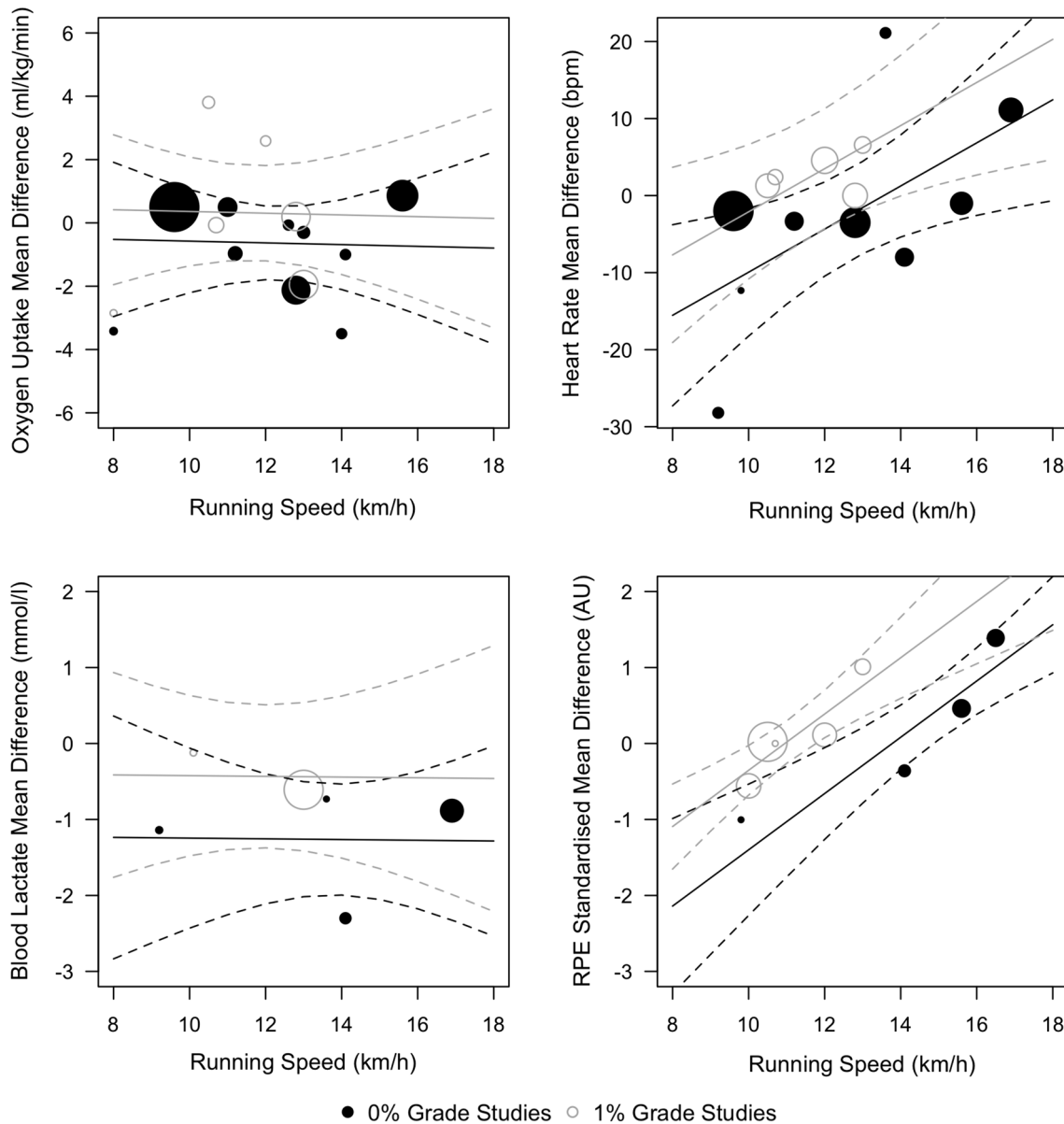


Fig. 3 Random-effects meta-regression of submaximal oxygen uptake (top left), heart rate (top right), blood lactate (bottom left) and rating of perceived exertion (bottom right) during treadmill compared with overground running based on running speed. Subgroup analysis based on relative treadmill grade. Positive values indicate outcome measure is higher for treadmill running. Larger datapoints received greater

weighting than smaller datapoints. Solid black datapoints indicate 0% treadmill grade. Open grey datapoints indicate 1% treadmill grade. Solid lines represent the estimated relationship and dashed lines represent the upper and lower limits of the 95% confidence interval. *RPE* rating of perceived exertion

3.7 Heart Rate

3.7.1 Submaximal Running

Pooled results indicated that submaximal motorised treadmill running at 0% relative grade reduced HR by 3 bpm compared with overground running (Fig. 5). This reduction was not significant ($p = 0.546$; $n = 122$; $k = 8$) and was affected by high heterogeneity ($I^2 = 97\%$). Pooled results indicated

that submaximal motorised treadmill running at 1% relative grade increased HR by 3 bpm compared with overground running (Fig. 5). This increase was significant ($p = 0.019$; $n = 38$; $k = 4$) but was affected by moderate heterogeneity ($I^2 = 31\%$). The GRADE quality of evidence was moderate for the overall motorised treadmill pooled results (Table 3). Motorised treadmill running speed was a significant moderator of the MD in submaximal HR ($\beta = 3$ bpm; 95% CI 1–5; $p = 0.012$; Fig. 3). HR was estimated to be similar between

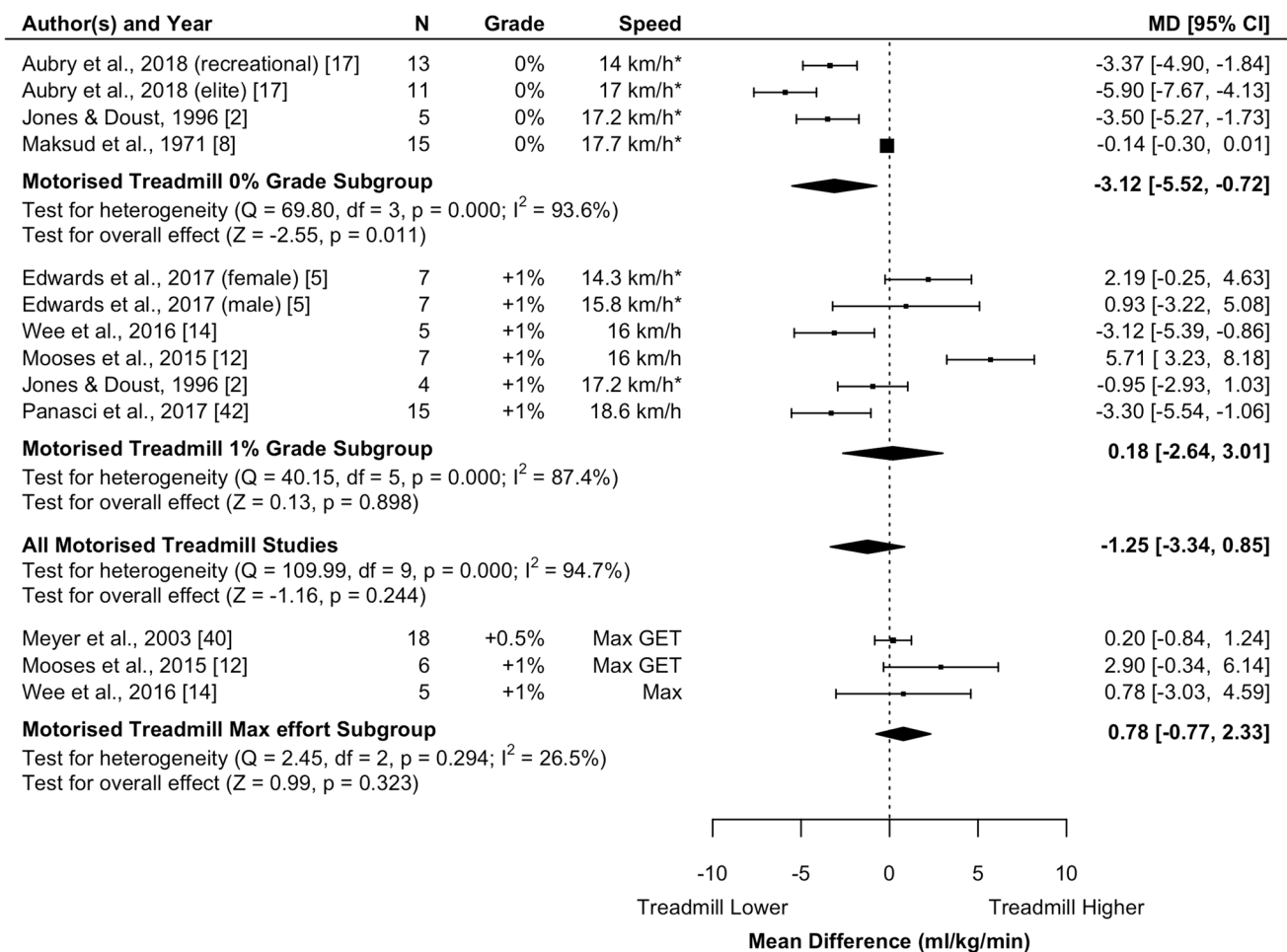


Fig. 4 Random-effects meta-analysis of near-maximal ($\geq 80\%$ $\dot{V}O_{2max}$) and maximal oxygen uptake during treadmill compared with overground running. Subgroup analysis based on relative treadmill

grade. *Mean submaximal running speed used in study. *CI* confidence interval, *df* degrees of freedom, *GET* graded exercise test, *Max* maximum, *MD* mean difference, *N* sample size, $\dot{V}O_{2max}$ maximal oxygen uptake

overground and motorised treadmill running at 13.6 km/h for 0% grade and 10.8 km/h for 1% grade.

One study investigating the effect of treadmill grades ranging from 0% to 3% in nine runners indicated that 1% grade resulted in the most closely matched HR between submaximal motorised treadmill and overground running [2].

3.7.2 Maximal Running

Pooled results indicated that near-maximal ($\geq 80\%$ $\dot{V}O_{2max}$) motorised treadmill running at 0% grade significantly reduced HR by 3 bpm ($p = 0.011$; $n = 20$; $k = 2$) compared with overground running but there was no significant difference at 1% grade ($p = 0.633$; $n = 45$; $k = 5$) (Fig. 6). These pooled differences were affected by no ($I^2 = 0\%$) or moderate ($I^2 = 47\%$) heterogeneity, respectively (Fig. 6). The GRADE quality of evidence was moderate for the 0% and 1% subgroups and low for the overall motorised treadmill pooled

results (Table 3). Pooled results indicated no significant difference in HR between motorised treadmill and overground running at maximal intensity ($p = 0.518$; $n = 23$; $k = 2$) with low heterogeneity ($I^2 = 18\%$; Fig. 6) and moderate GRADE quality of evidence (Table 3).

3.8 Lactate

3.8.1 Submaximal Running

Pooled results indicated that submaximal motorised treadmill running reduced La compared to overground running by 1.26 and 0.52 mmol/L when the treadmill was set to 0% and 1% relative grade, respectively (Fig. 7a). These differences were significant at 0% ($p < 0.001$; $n = 37$; $k = 3$) and 1% relative grade ($p = 0.041$; $n = 16$; $k = 2$). There was no heterogeneity ($I^2 = 0.0\%$) for the pooled estimate at 1% treadmill grade and moderate heterogeneity ($I^2 = 48\%$) for

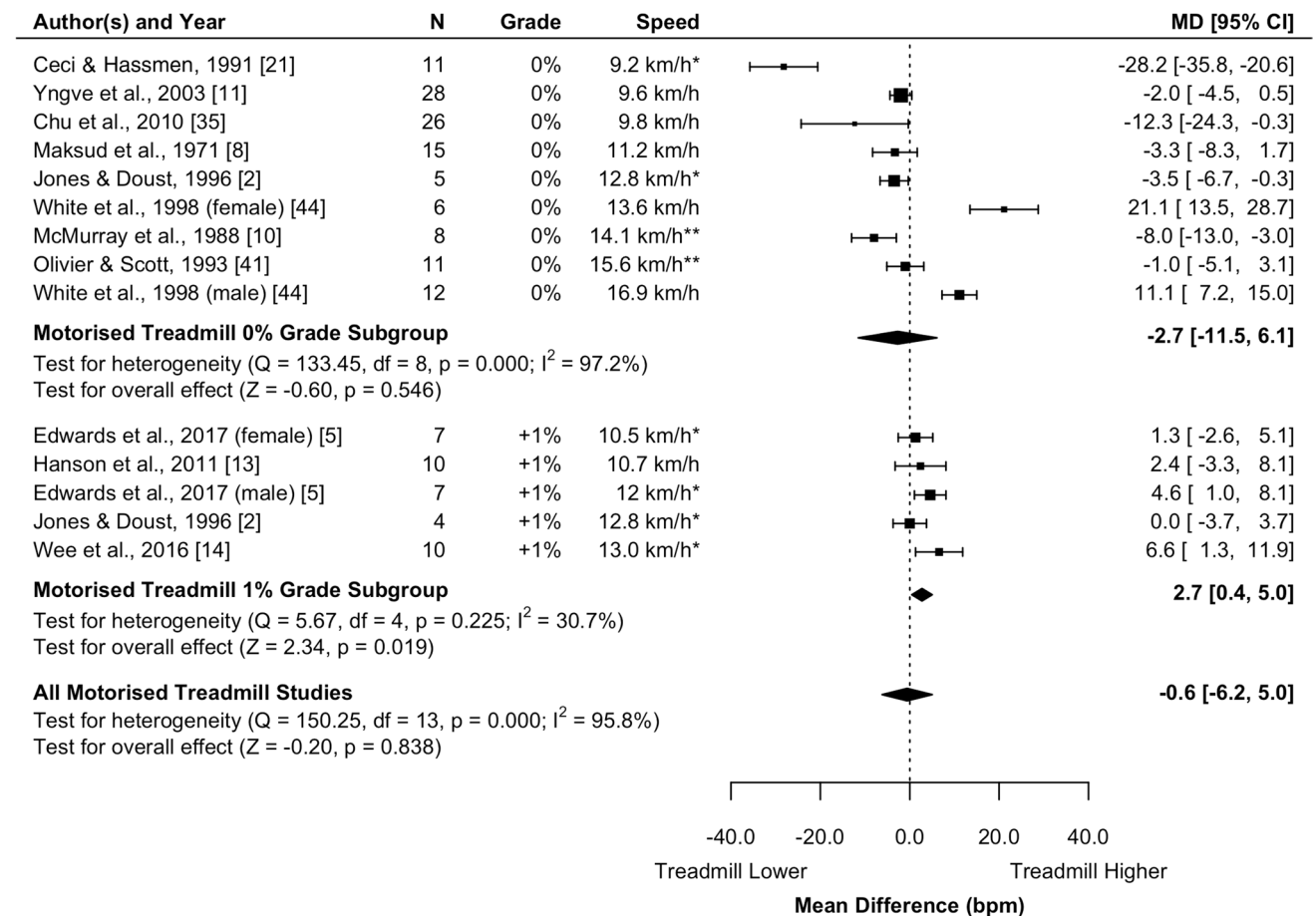


Fig. 5 Random-effects meta-analysis of submaximal heart rate during treadmill compared with overground running. Subgroup analysis based on relative treadmill grade. *Mean submaximal running speed

used in study. **Running speed estimated from the Jones and Doust [2] study that involved runners of similar ability. *CI* confidence interval, *df* degrees of freedom, *MD* mean difference, *N* sample size

pooled results at 0% treadmill grade. The GRADE quality of evidence was moderate for the 1% grade subgroup and low for the 0% grade subgroup and overall motorised treadmill pooled results (Table 3). Motorised treadmill running speed was not a significant moderator of the MD in submaximal La ($\beta = 0.00$ mmol/L; 95% CI -0.25 to 0.24; $p = 0.970$; Fig. 3).

3.8.2 Maximal Running

Pooled results indicated that near-maximal ($\geq 80\%$ $\dot{V}O_{2\max}$) motorised treadmill running reduced blood lactate by 0.54 mmol/L compared with overground running (Fig. 7b). This difference was not significant ($p = 0.092$; $n = 20$; $k = 2$), was affected by low heterogeneity ($I^2 = 0\%$) and provided moderate GRADE quality of evidence (Table 3).

3.9 Rating of Perceived Exertion

3.9.1 Submaximal Running

Pooled results indicated that submaximal motorised treadmill running caused a small, non-significant increase in RPE compared with overground running when the treadmill was set to 0% ($p = 0.714$; $n = 55$; $k = 4$) and a trivial non-significant increase for 1% relative grade ($p = 0.702$; $n = 56$; $k = 4$) (Fig. 8a). Both pooled results were affected by high heterogeneity ($I^2 > 77\%$). The GRADE quality of evidence was low for the 1% grade subgroup and very low for the 0% grade subgroup and overall motorised treadmill pooled results (Table 3). Motorised treadmill running speed was a significant moderator of the MD in submaximal RPE (SMD $\beta = 0.37$; 95% CI 0.21–0.53; $p < 0.001$; Fig. 3). RPE was estimated to be equal between overground and motorised

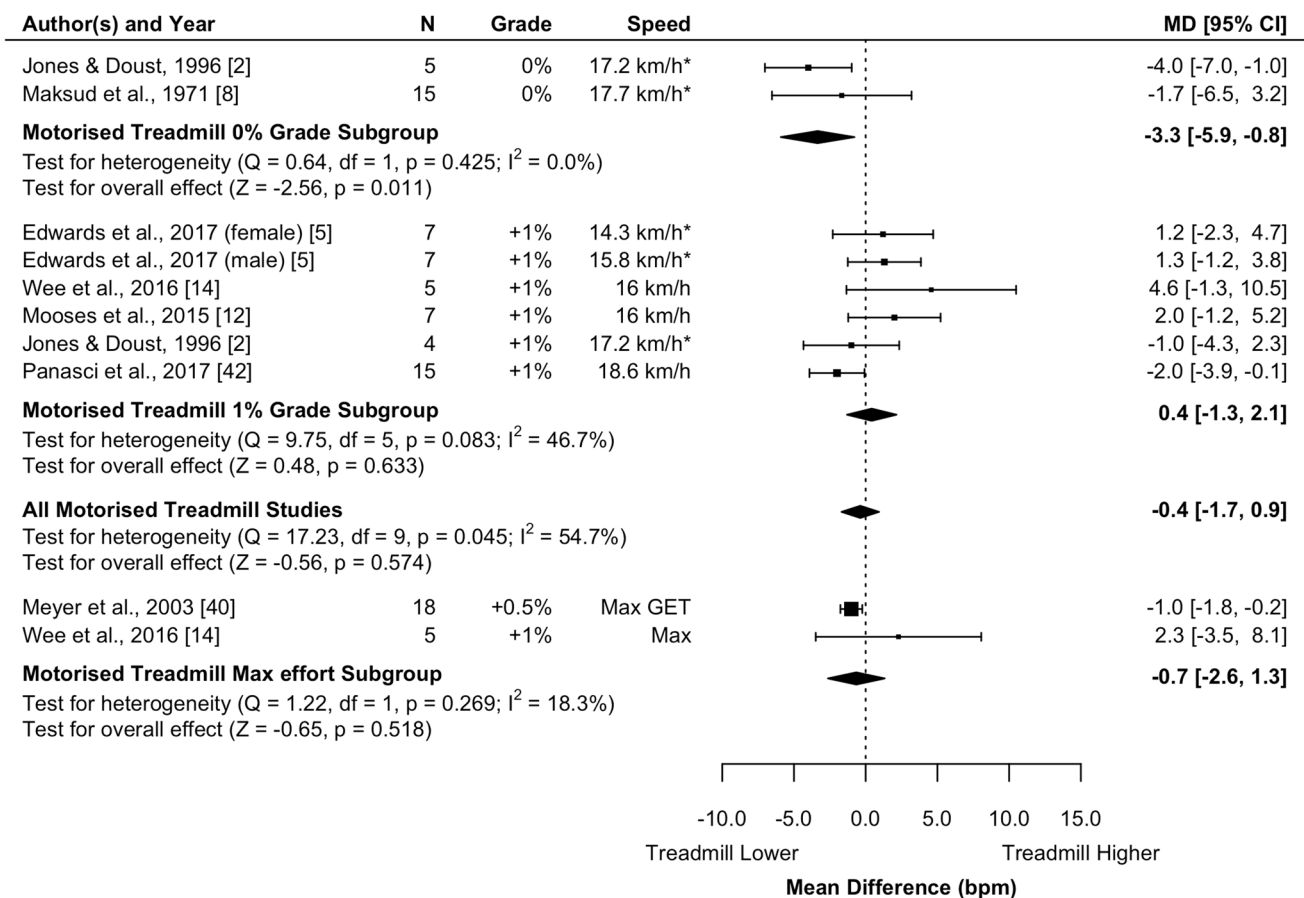


Fig. 6 Random-effects meta-analysis of near-maximal ($\geq 80\% \dot{V}O_{2max}$) and maximal heart rate during treadmill compared with overground running. Subgroup analysis based on relative treadmill grade.

*Mean submaximal running speed used in study. *CI* confidence interval, *df* degrees of freedom, *GET* graded exercise test, *MD* mean difference, *N* sample size, $\dot{V}O_{2max}$ maximal oxygen uptake

treadmill running at 13.8 km/h for 0% grade and 11.0 km/h for 1% grade.

3.9.2 Maximal Running

Pooled results indicated that RPE during near-maximal ($\geq 80\% \dot{V}O_{2max}$) motorised treadmill running at 1% grade was not significantly different from overground ($p = 0.672$; $n = 39$; $k = 3$) (Fig. 8b). This pooled result was affected by large heterogeneity ($I^2 = 95\%$) and provided low GRADE quality of evidence (Table 3).

RPE during a 5 km time trial on a treadmill with 1% grade was significantly greater than overground running indoors (SMD = 0.42) or outdoors (SMD = 0.76) [39].

3.10 Submaximal Velocity

One study compared the submaximal running speeds selected by runners when they were asked to run at three

different fixed perceived levels of effort on a motorised treadmill and overground [21]. The runners chose significantly slower running speeds when running on the motorised treadmill than overground (MD 3.2–5.9 km/h across speeds) [21].

3.11 Endurance Performance

3.11.1 Pooled Results

Pooled results indicated that performing endurance time trial testing on a treadmill with 0% relative grade was associated with small, significant decreases in 5 km performance (i.e. increased time taken to run a 5 km distance) when compared with overground ($p < 0.001$; $n = 18$; $k = 2$) (Fig. 9a). This difference was equivalent to ~ 40 s over 5 km. This result was not affected by any heterogeneity ($I^2 = 0.0\%$) and provided low GRADE quality of evidence (Table 3). Pooled results indicated that performing endurance testing on a

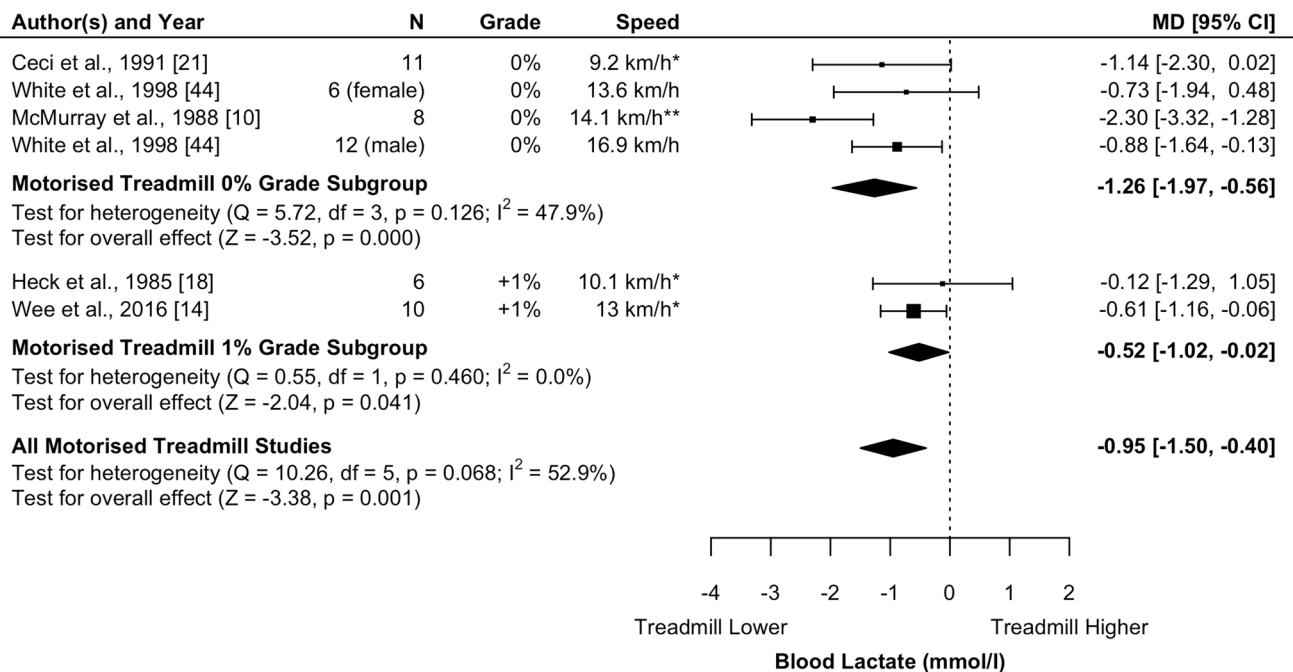
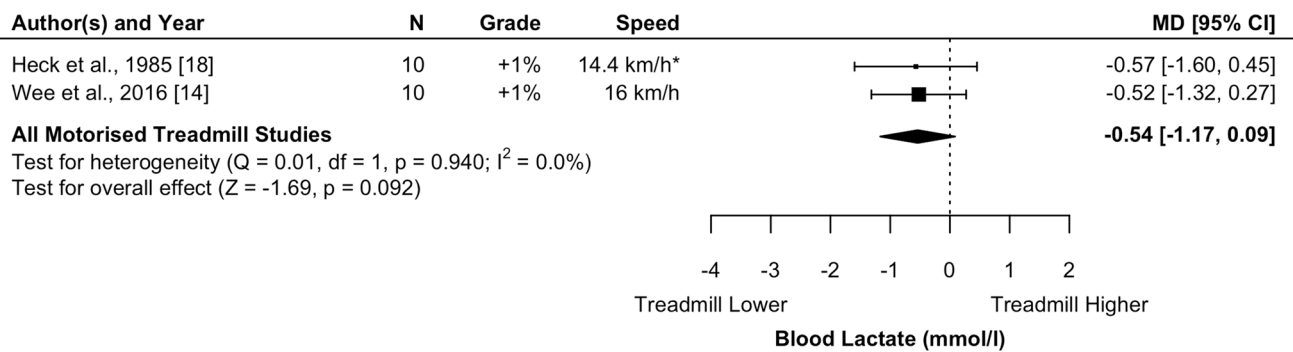
a Blood lactate concentration (submaximal speeds)**b Blood lactate concentration (near-maximal speeds)**

Fig. 7 Random-effects meta-analysis comparing blood lactate during treadmill and overground running at submaximal (a) and near-maximal ($\geq 80\% \dot{V}O_{2max}$) (b) running intensity. Subgroup analysis based on relative treadmill grade. *Mean submaximal running speed used

in study. **Running speed estimated from the Jones and Doust [2] study that involved runners of similar ability. CI confidence interval, df degrees of freedom, MD mean difference, N sample size, $\dot{V}O_{2max}$ maximal oxygen uptake

treadmill with 1% relative grade was associated with small, non-significant decreases in performance (i.e. increased time taken to complete a 5–10 km set distance or decreased distance covered in a 1 h trial) when compared with overground ($p = 0.234$; $n = 88$; $k = 3$) (Fig. 9a). This result was affected by large heterogeneity ($I^2 = 96\%$) and provided very low GRADE quality of evidence (Table 3).

Results from one study indicated that maximal aerobic speed determined on a treadmill was 2.0 km/h (95% CI 1.3–2.7) faster than testing on an overground track [1].

3.12 Sprint Performance

3.12.1 Pooled Results

Sprint performance results varied considerably across studies (Fig. 9b). Of the three studies identified, one indicated no difference in 91 m sprint maximum speed amongst five runners (MD: 0.0 km/h) [19], one indicated a 7.0 km/h reduction in 100 m sprint maximum speed amongst non-runners [6] and one indicated a 3.0 km/h increase in 10 s

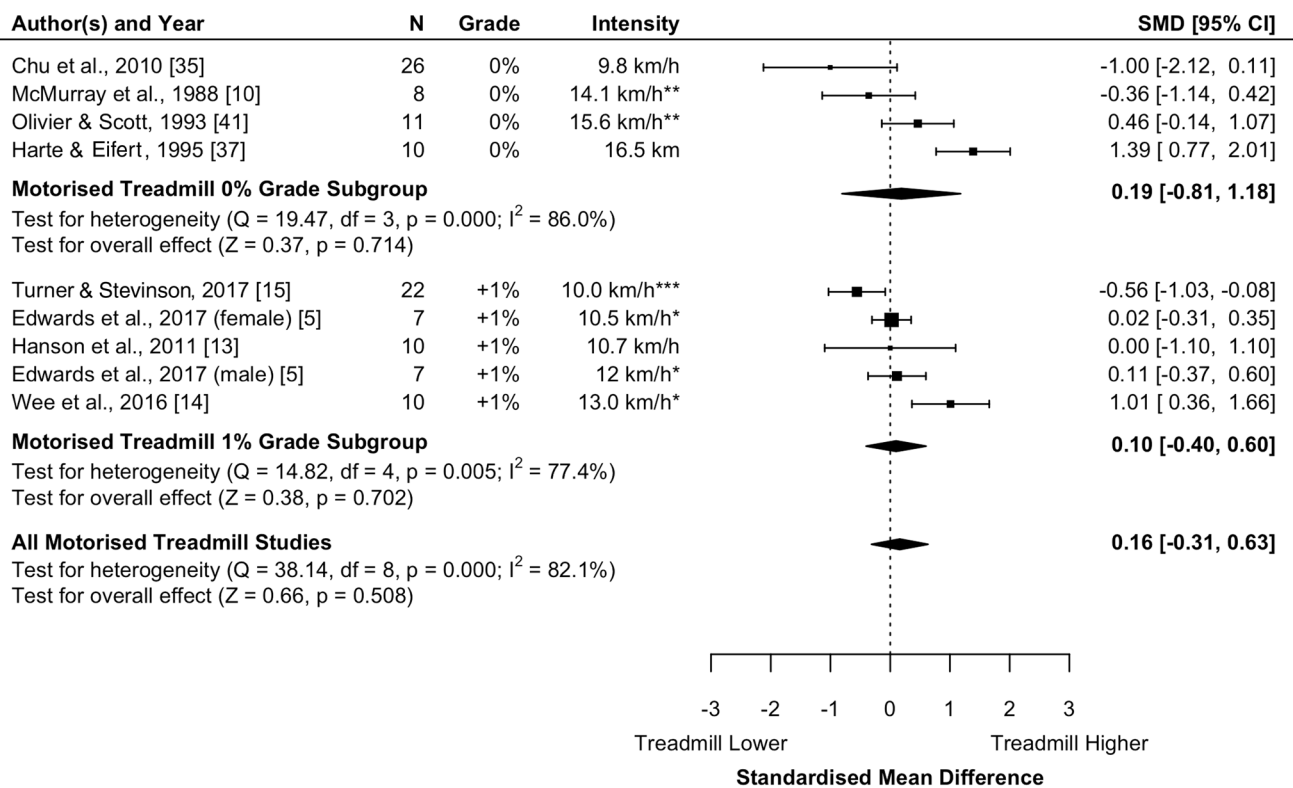
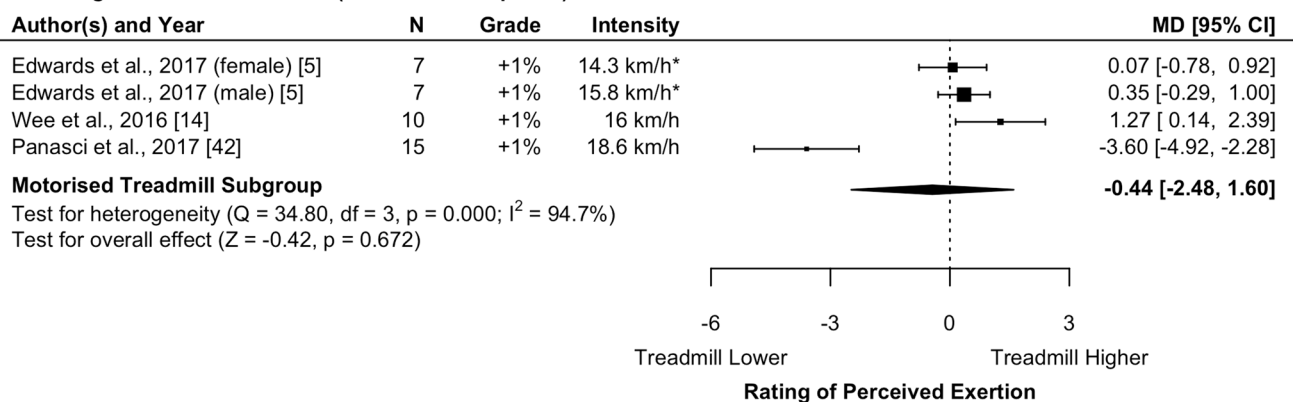
a Rating of Perceived Exertion (submaximal speeds)**b Rating of Perceived Exertion (near-maximal speeds)**

Fig. 8 Random-effects meta-analysis comparing rating of perceived exertion during treadmill and overground running at submaximal (**a**) and near-maximal ($\geq 80\% \dot{V}O_{2max}$) (**b**) running intensity. Subgroup analysis based on relative treadmill grade. *Mean submaximal running speed used in study. **Running speed estimated from the Jones

and Doust [2] study that involved runners of similar ability. ***Estimated based on running speed being 12.0 km/h during the maximal stage of testing. CI confidence interval, df degrees of freedom, N sample size, SMD standardised mean difference, $\dot{V}O_{2max}$ maximal oxygen uptake

sprint maximum speed amongst non-runners [33]. Pooled estimates indicated no significant difference ($p \geq 0.317$) and were affected by large heterogeneity (Fig. 9b), and provided very low GRADE quality of evidence (Table 3).

4 Discussion

The aim of this review was to compare physiological, perceptual and performance measures between treadmill and overground running. Thirty-four studies identified by the search were eligible for inclusion; all but one study considered motorised treadmill running, whereas only three studies

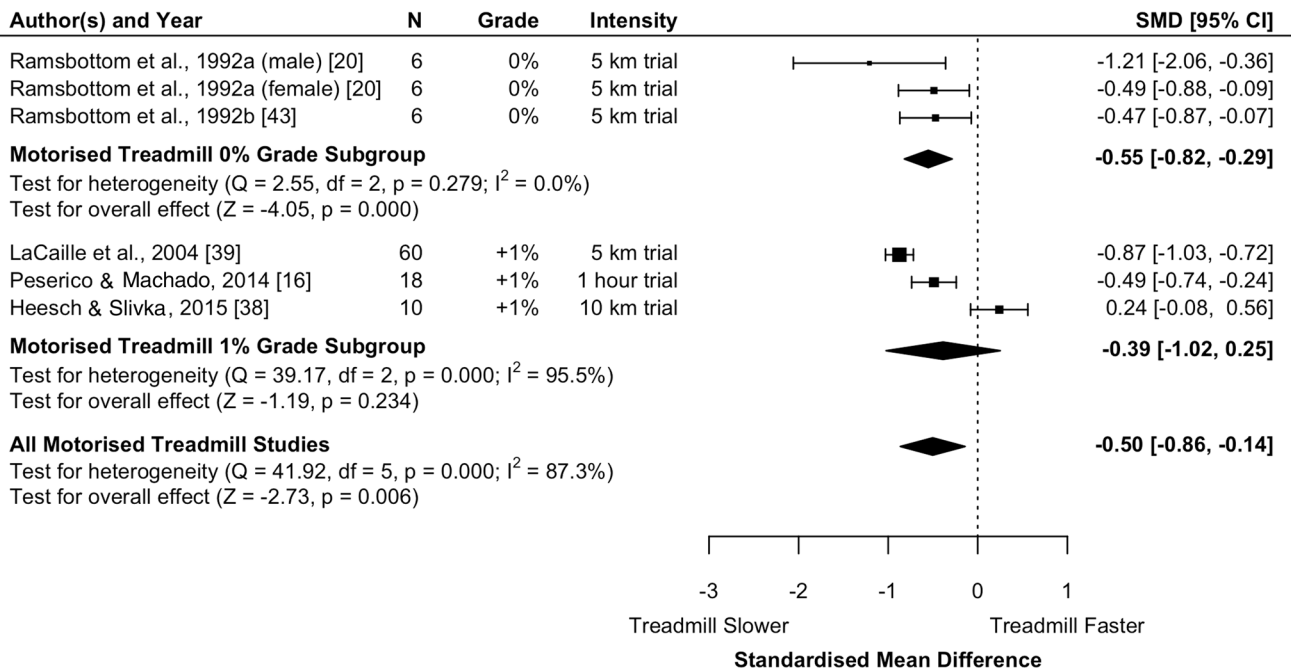
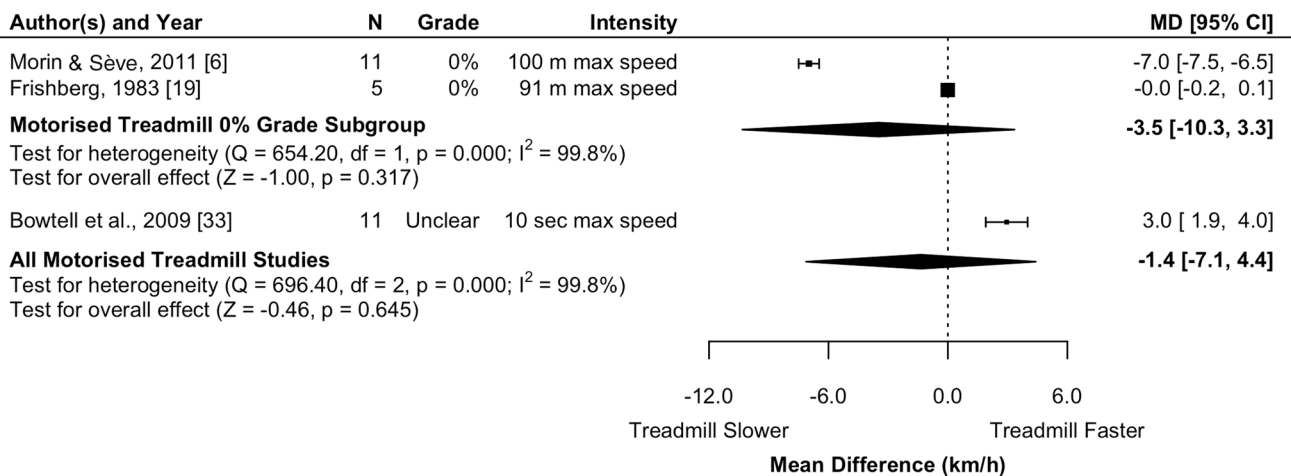
a Endurance Performance**b Sprint Performance**

Fig. 9 Random-effects meta-analysis of endurance (a) and sprint (b) performance during treadmill compared with overground running. Subgroup analysis based on relative treadmill grade. CI confidence

interval, *df* degrees of freedom, *max* maximum, *MD* mean difference, *N* sample size, *SMD* standardised mean difference

considered non-motorised treadmill running. Meta-analyses comparing submaximal motorised treadmill and overground running suggested that (1) $\dot{V}O_2$ was not different at 0% or 1% treadmill grade (low-to-moderate evidence); (2) La was lower at 0% and to a lesser extent 1% treadmill grade (low-to-moderate evidence); and (3) HR and RPE were higher than overground running at faster submaximal speeds but lower at slower speeds (very low-to-moderate evidence). Meta-analyses comparing near-maximal and maximal motorised treadmill and overground running suggested that

(1) near-maximal $\dot{V}O_2$ and HR were lower at 0% and similar at 1% treadmill grade (low evidence); (2) $\dot{V}O_{2max}$ and maximal HR were similar (moderate evidence); (3) La tended to be lower at 1% treadmill grade (moderate evidence); and (4) RPE tended to be similar (low evidence). Endurance running performance was better overground than motorised treadmill (low evidence), whereas sprint performance did not differ consistently (very low evidence).

To the authors' knowledge, this is the first attempt to systematically review and synthesise studies that have

compared treadmill and overground running. Summarising this literature is important because treadmills are commonly used in clinical and research settings to simulate the demands of overground running and a high percentage of runners regularly train on treadmills [45]. Any discrepancies between treadmill and overground running may lead to incorrect prescription of training workloads. This could lead to poor rehabilitation outcomes in clinical settings and suboptimal adaptations to training amongst competitive runners. Findings from this review allow practitioners to more precisely apply treadmill-based assessments of running to real-life conditions, which could improve rehabilitation and performance outcomes.

Submaximal $\dot{V}O_2$ was the most commonly used variable for comparing motorised treadmill and overground running. Meta-analysis revealed that there was no overall difference in submaximal $\dot{V}O_2$ between motorised treadmill and overground running, although there was considerable uncertainty across studies. CIs from the meta-analysis suggested that differences were between 1.12 mL/kg/min lower and 0.56 mL/kg/min higher for treadmill running, with values tending to be lower at 0% grade and higher at 1% treadmill grade. The null finding for submaximal $\dot{V}O_2$ is notable considering the widespread belief that a 1% treadmill grade effectively matches the physiological workload of submaximal overground running but a 0% treadmill grade does not. Jones and Doust [2] were the first to propose that a 1% treadmill grade is required to compensate for the lack of air resistance during treadmill compared with overground running. This has become widespread practice among researchers seeking to match workloads between motorised treadmill and overground running, with the paper cited over 600 times. However, the results of this review suggest that researchers, coaches and athletes can infer overground running economy at submaximal speeds $< 80\% \dot{V}O_{2max}$ (typically equivalent to $< 13\text{--}16$ km/h for runners included in the meta-analysis) from motorised treadmill $\dot{V}O_2$ testing without adjusting treadmill grade.

$\dot{V}O_2$ during near-maximal motorised treadmill running was lower at 0% grade but similar at 1% grade compared with overground running. This finding provides some support for following the Jones and Doust [2] 1% treadmill grade recommendation at higher running intensities ($\geq 80\% \dot{V}O_{2max}$) and is consistent with the energy required to overcome air resistance becoming greater at faster running speeds [46]. Two previous studies have also demonstrated greater $\dot{V}O_2$ differences between motorised treadmill and overground running at near-maximal compared with submaximal speeds [8, 17]. $\dot{V}O_2$ findings at near-maximal speeds should be interpreted with caution because they require greater involvement of anaerobic processes and do not allow energy expenditure to be inferred from $\dot{V}O_2$. Nonetheless, our findings suggest that treadmill-based measurements of the running velocities

associated with $> 80\% \dot{V}O_{2max}$ are likely to overestimate the equivalent overground running speed. A 1% treadmill grade should be used if accurate determination of running velocities associated with $> 80\% \dot{V}O_{2max}$ are required (typically equivalent to $> 13\text{--}16$ km/h for runners included in the meta-analysis). There was no difference between $\dot{V}O_{2max}$ values measured on a motorised treadmill compared with overground; therefore, researchers, coaches and athletes can rely on $\dot{V}O_{2max}$ treadmill testing as a valid measure of aerobic power.

La values were lower during motorised treadmill running than overground running, irrespective of running speed. La values were more closely matched between conditions at 1% compared with 0% treadmill grade; however, values remained lower for the motorised treadmill condition, suggesting that a 1% treadmill grade may not be enough to match La values compared with overground running. This explanation is not consistent with other findings of this review that suggest submaximal $\dot{V}O_2$ is similar between motorised treadmill and overground running. The lactate meta-analysis was based on fewer studies (5 vs. 12 studies) and fewer total participants (53 vs. 164 participants) than the submaximal $\dot{V}O_2$ meta-analysis, so the conflicting results may relate to the lower precision of the analysis. Additionally, submaximal blood lactate measurements are associated with high typical error (approximately 27%) [47] and this could have influenced the accuracy of individual studies included in the lactate meta-analysis. Alternatively, La values and the associated anaerobic processes might be more sensitive than $\dot{V}O_2$ to the proposed greater energy demands of overground running. Nonetheless, findings from this review suggest that a greater than 1% grade is needed to accurately match La values between treadmill and overground running. This finding is important because coaches commonly use running velocities associated with different La values to prescribe training intensities.

HR and RPE tended to be similar between motorised treadmill and overground running at submaximal speeds but speed moderated the direction and magnitude of the difference. Overground and motorised treadmill running were best matched at 14 and 11 km/h for 0% and 1% treadmill grade, respectively. HR and RPE were lower for motorised treadmill running if submaximal speeds were below these values and higher if submaximal speeds were above these values. The higher HR and RPE during motorised treadmill than during overground running at faster submaximal speeds was unexpected, especially since air resistance increases with an increase in speed [46]. The difference might relate to participants feeling less comfortable when running at faster speeds on a motorised treadmill. Treadmill running experience and comfort was rarely considered by studies included in this review. Most runners prefer running overground over

running on a treadmill [45] and we speculate this could partly relate to having greater control over their speed and ability to stop moving, and a lower risk of falling. If correct, the lack of comfort with higher-speed treadmill running might cause greater increases in HR and perceived effort due to feelings of anxiety about falling. The greater perceived effort might also relate to greater increases in body temperature during treadmill running because there is less relative air movement that can reduce body temperature due to evaporation of sweat [44]. Consistent with these explanations, one study included in this review [21] and another study that was excluded from this review due to participants being adolescents [48] demonstrated that participants selected slower running speeds on a motorised treadmill than overground when asked to exercise at fixed levels of perceived effort. Collectively, these findings suggest that researchers, coaches and athletes should be cautious when attempting to apply HR and RPE measures derived from submaximal treadmill running to overground running.

Based on the aforementioned relationship between submaximal speed, HR and RPE, large differences in HR and RPE between motorised treadmill and overground running would be expected at near-maximal and maximal intensities. However, HR and RPE were similar or slightly lower at near-maximal and maximal intensities. We speculate that as running speed approaches maximum values, the differences in HR and RPE between treadmill and overground running will be minimal because runners will reach their bodily limits. For example, we do not expect that fear of falling on a treadmill can increase HR beyond physiological limits. As a result, participant comfort and experience with treadmill running is likely to be most important when assessing HR and RPE at submaximal speeds.

Pooled results provided low evidence that running performance during set-distance or set-duration trials on a motorised treadmill was poorer than equivalent trials performed overground. This finding is contradictory to the lower $\dot{V}O_2$ values and similar $\dot{V}O_2$ during treadmill running; however, it is consistent with the higher RPE during treadmill running at faster speeds and the observation that runners choose to run at slower speeds on treadmills than overground. It is likely that participants feel less comfortable during maximal effort running on treadmills than overground (due to falling risks and less effective thermoregulation) and therefore select slower pacing strategies, despite the reduced air resistance that needs to be overcome during treadmill running. Additionally, treadmill running is absent of the extra attentional requirements inherent with overground running, allowing for increased internal focus on physiological sensations such as muscle fatigue or HR, which might further affect pacing strategy [38] or running technique and running economy [49].

Sprint performance varied considerably between motorised treadmill and overground running. Based on the meta-analysis CI, maximum sprint speed ranges from 7.1 km/h slower to 4.4 km/h faster when running on a motorised treadmill than overground. The considerable uncertainty across studies might reflect difficulties in replicating the acceleration and deceleration phases of sprinting on a treadmill more so than true differences in sprint performance between the two running mediums [19].

Several important limitations should be considered when interpreting the findings of this review. First, common methodological issues were found in the included studies, such as using a fixed order of conditions and difficulties in matching or controlling running speeds and environmental conditions. Most overground trials were conducted outdoors and were highly likely to be affected by variability in temperature, wind and humidity. Second, very few studies included treadmill running experience and comfort as a study eligibility criterion, despite familiarity and comfort with treadmill running being highly likely to influence study findings. As a result, this review was unable to explore whether treadmill running experience contributed to the inconsistent results across studies. Third, a high level of heterogeneity was present for most meta-analyses in this review and as such results should be interpreted with caution. Improved reporting of studies and open data will facilitate further exploration of possible moderating variables that may explain the high heterogeneity observed across studies. We attempted to account for some of the missing information by estimating values from other studies, but this approach is likely to introduce some error that could be avoided by improved reporting. Fourth, the high heterogeneity could also be caused by differences in the motorised treadmill models used across studies and this factor was not investigated in this review. Previous research has demonstrated a 7% difference in $\dot{V}O_2$ and 5% difference in HR when running at submaximal speed on motorised treadmills from different manufacturers [50]. $\dot{V}O_2$ and HR were higher for the treadmill with a stiffer treadmill platform and this should be considered by coaches and clinicians performing treadmill testing. Last, thermoregulation outcome measures were not included as review outcomes, but should be considered when attempting to match treadmill and overground running.

5 Conclusion

This review found that (1) $\dot{V}O_2$ was similar but LA values were lower and preferred running speeds were slower during submaximal running on a motorised treadmill than overground; (2) HR and RPE during submaximal treadmill running were higher at faster submaximal speeds and lower

at slower submaximal speeds than overground running; (3) $\dot{V}O_2$ and La values tended to be lower, whereas HR and RPE were similar during near-maximal running on a motorised treadmill compared with overground; (4) $\dot{V}O_{2\max}$ and maximal HR were similar across conditions; and (5) performance in set-distance and set-duration trials was poorer when performed on a motorised treadmill than overground but sprint performance varied considerably across conditions. As a result, it might not be possible to balance all outcomes between motorised treadmill and overground running. Researchers, clinicians, coaches and athletes should carefully select motorised treadmill speeds and grades based on the outcome measures they want to most closely match and the experience and comfort of the athlete with treadmill running.

Compliance with Ethical Standards

Funding Bas Van Hooren was funded by the Kootstra Talent Fellowship awarded by the Centre for Research Innovation, Support and Policy (CRISP) of Maastricht University Medical Center+. This research received no other specific grant from any funding agency.

Conflict of Interest Joel T. Fuller and Jonathan D. Buckley have been authors on some research projects that have evaluated the effects of different running shoes on running performance, biomechanics and physiology; those projects involved the use of running shoes that were donated by the shoe industry, either from running shoe retail stores or ASICS Oceania. Chris Bishop has received funding from both ASICS Oceania and Brittain Wynyard for professional services related to footwear. No companies played any role in the design, conduct or interpretation of the present research. Jayme R. Miller, Bas Van Hooren and Richard W. Willy declare no conflicts of interest.

Author Contributions All authors contributed to the conception and design of the review and completion of the search strategy. Joel T. Fuller was responsible for the meta-analysis and meta-regression. Jayme R. Miller drafted the manuscript. All authors edited and revised the manuscript and approved the final version of the manuscript.


Data Availability Statement The datasets generated and/or analysed during the current systematic review are available in Appendix S2 of the Electronic Supplementary Material.

References

- Cappa DF, García GC, Secchi JD, Maddigan ME. The relationship between an athlete's maximal aerobic speed determined in a laboratory and their final speed reached during a field test. *J Sports Med Phys Fit.* 2014;54(4):424–31.
- Jones AM, Doust JH. A 1% treadmill grade most accurately reflects the energetic cost of outdoor running. *J Sports Sci.* 1996;14(4):321–7.
- Bishop C, Hillier S, Thewlis D. The reliability of the Adelaide in-shoe foot model. *Gait Posture.* 2017;56:1–7.
- Sirotic A, Coutts AJ. The reliability of physiological and performance measures during simulated team-sport running on a non-motorised treadmill. *J Sci Med Sport.* 2008;11(5):500–9.
- Edwards RB, Tofari PJ, Cormack SJ, Whyte DG. Non-motorized treadmill running is associated with higher cardiometabolic demands compared with overground and motorized treadmill running. *Front Physiol.* 2017;8:914.
- Morin JB, Sève P. Sprint running performance: comparison between treadmill and field conditions. *Eur J Appl Physiol.* 2011;111(8):1695–703.
- Pugh LG. Oxygen intake in track and treadmill running with observations on the effect of air resistance. *J Physiol.* 1970;207(3):823–35.
- Maksud MG, Coutts KD, Hamilton LH. Time course of heart rate, ventilation, and $\dot{V}O_2$ during laboratory and field exercise. *J Appl Physiol.* 1971;30(4):536–9.
- McMiken DF, Daniels JT. Aerobic requirements and maximum aerobic power in treadmill and track running. *Med Sci Sports Exerc.* 1976;8(1):14–7.
- McMurray RG, Berry MJ, Vann RT, Hardy CJ, Sheps DS. The effect of running in an outdoor environment on plasma beta endorphins. *Ann Sports Med.* 1988;3(4):230–3.
- Yngve A, Nilsson A, Sjöström M, Ekelund U. Effect of monitor placement and of activity setting on the MTI accelerometer output. *Med Sci Sports Exerc.* 2003;35(2):320–6.
- Mooses M, Tippi B, Mooses K, Durussel J, Mäestu J. Better economy in field running than on the treadmill: evidence from high-level distance runners. *Biol Sport.* 2015;32(2):155–9.
- Hanson NJ, Berg K, Meendering JR, Ryan C. Oxygen cost of running barefoot vs. running shod. *Int J Sports Med.* 2011;32(6):401–6.
- Wee VM, Heimburg E, Tillaar R. Comparison of perceptual and physiological variables of running on a track, motorized treadmill, and non-motorized curved treadmill at increasing velocity. *Acta Kinesiol Univ Tartu.* 2016;22:20–35.
- Turner TL, Stevinson C. Affective outcomes during and after high-intensity exercise in outdoor green and indoor gym settings. *Int J Environ Health Res.* 2017;27(2):106–16.
- Peserico SC, Machado AF. Comparison between running performance in time trials on track and treadmill. *Braz J Kineanthrop Hum Perform.* 2014;16(4):456–64.
- Aubry RI, Power GA, Burr JF. An assessment of running power as a training metric for elite and recreational runners. *J Strength Cond Res.* 2018;32(8):2258–64.
- Heck H, Mader A, Hess G, Mucke S, Muller R, Hofmann W. Justification of the 4-mmol/l lactate threshold. *Int J Sports Med.* 1985;6:117–30.
- Frishberg BA. An analysis of overground and treadmill sprinting. *Med Sci Sports Exerc.* 1983;15(6):478–85.
- Ramsbottom R, Williams C, Kerwin DG, Nute ML. Physiological and metabolic responses of men and women to a 5-km treadmill time trial. *J Sports Sci.* 1992;10(2):119–29.
- Ceci R, Hassmén P. Self-monitored exercise at three different RPE intensities in treadmill vs field running. *Med Sci Sports Exerc.* 1991;23(6):732–8.
- Higgins JPT, Altman DG, Gotzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ.* 2011;343:d5928.
- Fuller JT, Bellenger CR, Thewlis D, Tsiros MD, Buckley JD. The effect of footwear on running performance and running economy in distance runners. *Sports Med.* 2015;45(3):411–22.
- Elbourne DR, Altman DG, Higgins JP, et al. Meta-analysis involving cross-over trials: methodological issues. *Int J Epidemiol.* 2002;31(1):140–9.
- Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for *t*-tests and ANOVAs. *Front Psychol.* 2013;4:863.
- Hopkins WG. A new view of statistics. <http://www.sportssci.org/resource/stats/>. Accessed 7 Sept 2018.

27. Higgins JP, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analysis. *BMJ*. 2003;327:557–60.
28. Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations. *BMJ*. 2004;328(7454):1490.
29. Stevens CJ, Hacene J, Wellham B, Sculley DV, Callister R, Taylor L, Dascombe BJ. The validity of endurance running performance on the curve 3TM non-motorised treadmill. *J Sports Sci*. 2015;33:1141–8.
30. Bassett DR Jr, Giese MD, Nagle FJ. Aerobic requirements of overground versus treadmill running. *Med Sci Sports Exerc*. 1985;17(4):477–81.
31. Bidder OR, Goulding C, van Walsum TA, Siebert U, Halsey LG. Does the treadmill support valid energetics estimates of field locomotion? *Integr Comp Biol*. 2017;57(2):301–19.
32. Bonen A, Gass GC, Kachadorian WA, Johnson RE. Energy cost of walking and running on different surfaces. *Aust J Sci Med Sport*. 1974;6(1):7.
33. Bowtell MV, Tan HL, Wilson AM. The consistency of maximum running speed measurements in humans using a feedback-controlled treadmill, and a comparison with maximum attainable speed during overground locomotion. *J Biomech*. 2009;42(15):2569–74.
34. Brookes FB, Knibbs AA, Pantlin CM, Wilson JK. An investigation into the biomechanical and physiological differences between road and treadmill running. *Res Pap Phys Educ*. 1971;2(2):28–35.
35. Chu CY, Lu SY, Lin KF. Influences of exercise experience and exercise settings on heart rate responses during self-selected intensity exercises. *J Exerc Sci Fit*. 2010;8(2):73–7.
36. Dal Monte A, Fucci S, Manoni A. The treadmill as a training and simulator instrument in middle and long distance running. *J Sports Med Phys Fit*. 1974;14(2):67–72.
37. Harte JL, Eifert GH. The effects of running, environment, and attentional focus on athletes catecholamine and cortisol-levels and mood. *Psychophysiology*. 1995;32(1):49–54.
38. Heesch MW, Slivka DR. Running performance, pace strategy, and thermoregulation differ between a treadmill and indoor track. *J Strength Cond Res*. 2015;29(2):330–5.
39. LaCaille RA, Masters KS, Heath EM. Effects of cognitive strategy and exercise setting on running performance, perceived exertion, affect, and satisfaction. *Psychol Sport Exerc*. 2004;5(4):461–76.
40. Meyer T, Welter JP, Scharhag J, Kindermann W. Maximal oxygen uptake during field running does not exceed that measured during treadmill exercise. *Eur J Appl Physiol*. 2003;88:387–9.
41. Olivier S, Scott PA. Physiological, perceptual and attitudinal responses to identically matched workloads in field and laboratory testing conditions. *S Afr J Res Sport Phys Educ Recreat*. 1993;16(1):63–72.
42. Panasci M, Lepers R, La Torre A, Bonato M, Assadi H. Physiological responses during intermittent running exercise differ between outdoor and treadmill running. *Appl Physiol Nutr Metab*. 2017;42(9):973–7.
43. Ramsbottom R, Colquhoun D, Williams C, Nute ML. Physiological and metabolic responses of trained runners to a 5 km treadmill time trial. *Aust J Sci Med Sport*. 1992;24(1):8–11.
44. White JA, Pomfret DK, Rennie S, Wong J, Ford M. Fluid replacement needs of well-trained male and female athletes during indoor and outdoor steady state running. *J Sci Med Sport*. 1998;1(3):131–42.
45. Taunton JE, Ryan MB, Clement DB, McKenzie DC, Lloyd-Smith DR, Zumbo BD. A prospective study of running injuries: the Vancouver Sun Run “in training” clinics. *Br J Sports Med*. 2003;37(3):239–44.
46. Davies CT. Effects of wind assistance and resistance on the forward motion of a runner. *J Appl Physiol Respir Exerc Physiol*. 1980;48(4):702–9.
47. Saunders PU, Pyne DB, Telford RD, Hawley JA. Reliability and variability of running economy in elite distance runners. *Med Sci Sports Exerc*. 2004;36(11):1972–6.
48. Kong PW, Koh TM, Tan WC, Wang YS. Unmatched perception of speed when running overground and on a treadmill. *Gait Posture*. 2012;36(1):46–8.
49. Schucker L, Parrington L. Thinking about your running movement makes you less efficient: attentional focus effects on running economy and kinematics. *J Sports Sci*. 2019;37(6):638–46.
50. Smith JAH, McKerrow AD, Kohn TA. Metabolic cost of running is greater on a treadmill with a stiffer running platform. *J Sports Sci*. 2017;35:1592–7.
51. Miller J, Van Hooren B, Bishop C, Buckley J, Willy R, Fuller J. A systematic review comparing physiological, perceptual and performance measures between treadmill and overground running. *OSF*. 2018. <https://doi.org/10.17605/OSF.IO/R7BK6>.

Affiliations

Jayne R. Miller¹ · Bas Van Hooren^{2,3}  · Chris Bishop^{1,4}  · Jonathan D. Buckley¹  · Richard W. Willy⁵  · Joel T. Fuller⁶ 

¹ Alliance for Research in Exercise, Nutrition and Activity (ARENA), School of Health Sciences, University of South Australia, Adelaide, SA 5001, Australia

² Department of Nutrition and Movement Sciences, NUTRIM School of Nutrition and Translational Research in Metabolism, Maastricht University Medical Centre+, Maastricht, The Netherlands

³ Institute of Sport Studies, Fontys University of Applied Sciences, Eindhoven, The Netherlands

⁴ The Biomechanics Lab, Adelaide, SA, Australia

⁵ School of Physical Therapy and Rehabilitation Sciences, University of Montana, Missoula, MT, USA

⁶ Department of Health Professions, Faculty of Medicine and Health Sciences, Macquarie University, 75 Talavera Rd, Macquarie Park, NSW 2109, Australia