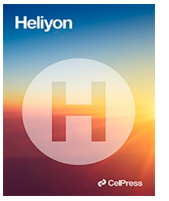
Heliyon 10 (2024) e39648

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Diagnostic accuracy of the lactate stress test for detecting mitochondrial disorders: Systematic review and meta-analysis

Sara El Guessabi [a,*], Redouane Abouqal [b], Azeddine Ibrahimi [c], Ghizlane Zouiri [d], Fatima Sfifou [e], Josef Finsterer [f], Yamna Kriouile [a,d]

[a] *Laboratory of genomics and Molecular Epidemiology of genetic Diseases: genes and Mutations in the Moroccan Population, Rabat Medical & Pharmacy School, Mohammed V University in Rabat, 6203, Morocco*
[b] *Laboratory of Biostatistics, Clinical Research and Epidemiology, Public Health Department, Rabat Medical & Pharmacy School, Mohammed V University in Rabat, 6203, Morocco*
[c] *Biotechnology Lab (MedBiotech), Bioinova Research Center, Rabat Medical & Pharmacy School, Mohammed V University in Rabat, 6203, Morocco*
[d] *Department of Paediatrics 2, Unit of Endocrinology and Neuropediatrics, Children's Hospital of Rabat, 6527, Morocco*
[e] *Laboratory of Histology Embryology and Cytogenetics, Rabat Medical & Pharmacy School, Mohammed V University in Rabat, 6203, Morocco*
[f] *Neurology Department, Neurology & Neurophysiology Center, Postfach 20, 1180, Vienna, Austria*

ARTICLE INFO

Keywords:
Mitochondrial disorders
Absolute load lactate stress test
Relative load lactate stress test
Diagnostic accuracy
Meta-analysis

ABSTRACT

Due to their variable phenotypes, mitochondrial disorders (MDs) can be difficult to diagnose. The absolute load lactate stress test (LSTA) and the relative load lactate stress test (LSTR) have been shown to be useful screening tools for the detection of MDs. In this study, we aimed to perform a meta-analysis to evaluate the diagnostic accuracy of these tests in detecting MDs. The study protocol was registered with PROSPERO (no. CRD42022331710). We performed a comprehensive search of PubMed, Web of Science and Scopus from January 10th, 2022 to July 27th, 2022 and included case-control and cohort diagnostic studies that targeted participants with MDs and used LSTA and/or LSTR as index tests. Two reviewers worked separately to compile information from selected articles. Risk of bias and applicability were assessed using the QUADAS-2 tool. Sensitivity and specificity, as well as diagnostic odds ratios (DORs) and area under the curve (AUC) were calculated using Meta-DiSc 2.0 and Stata software. The analysis included 14 studies with a total of 1064 participants, divided into six studies with 793 participants for LSTA and eight studies with 271 participants for LSTR. For LSTA the meta-analysis gave a pooled sensitivity of 0.67 (95 % CI 0.62, 0.72), a specificity of 0.93 (95 % CI 0.85, 0.97), DOR of 26.63 (95 % CI 10.99, 64.52), and AUC of 0.70 (95 % CI 0.66, 0.74). For LSTR, the pooled sensitivity was 0.52 (95 % CI 0.33, 0.70), specificity 0.94 (95 % CI 0.79, 0.99), DOR 18.14 (95 % CI 2.99, 109.85), and the AUC 0.80 (95 % CI 0.76, 0.83). LSTA and LSTR showed as screening tests moderate sensitivity and high specificity for MD diagnosis, particularly for LSTR. The choice of test may depend on the patient's individual aerobic capacity and motor skills and the availability of equipment.

* Corresponding author. Laboratory of genomics and molecular epidemiology of genetic diseases: genes and mutations in the Moroccan population, Medical & Pharmacy School, Mohammed V University in Rabat, Mohamed Belarbi El Alaoui street, BP 6203, Rabat Institut, Rabat, Morocco.
E-mail address: sara_elguessabi@um5.ac.ma (S. El Guessabi).

https://doi.org/10.1016/j.heliyon.2024.e39648
Received 8 June 2024; Received in revised form 12 September 2024; Accepted 21 October 2024
Available online 22 October 2024

1. Introduction

Mitochondrial disorders (MDs) are inborn errors of metabolism resulting from disturbances in cellular energy production and have an estimated prevalence of 1:5000 [1]. MDs present with a wide range of symptoms and signs and can affect any organ or tissue in the body [2]. The symptoms reported by MD patients are very diverse and vary with age at onset and even between individuals within the same family. This complexity makes it difficult to understand, diagnose and effectively treat these conditions. Clinical presentations include a wide range of symptoms and signs such as fatigue, muscle weakness, exercise intolerance, spasticity, dystonia, hypotonia, myalgia, short stature, developmental delay, stroke-like episodes, seizures, intellectual disability, learning difficulties, ophthalmoparesis, visual impairment, nystagmus, hearing loss, hypoglycemia, diabetes, neuropsychiatric symptoms, as well as complications involving the endocrine, digestive, kidney, heart, and liver systems [3].

The diagnosis of MDs is challenging and requires a wide range of techniques including assessment of clinical presentation, biomarkers, measurement of enzymatic activities of mitochondrial respiratory chain complexes in isolated muscle biopsy specimens, and detection of genetic mutations in nuclear DNA (nDNA) or mitochondrial DNA (mtDNA) by next-generation sequencing (NGS) [4]. In patients with suspected MD, the lactate stress test (LST) can be a valuable screening method alongside other diagnostic tests. For the diagnosis of MDs, it is important to use a combination of tests, including LST, as part of a comprehensive diagnostic approach [5].

The principle of the LST lies in its ability to more reliably detect disturbances in muscle energy metabolism during light exercise that may not be visible at rest but only under physical stress [6–9]. The LST is performed to measure the aerobic capacity of mitochondria [10].

Due to defects in respiratory chain enzymes in mitochondria, there is an increase in serum lactate during exercise due to a shift from aerobic to anaerobic phosphorylation of ADP, resulting in a shift toward lactate production in the pyruvate/lactate reaction. There is an increase in pyruvate levels and the ratio of NADH to NAD in the mitochondria also increases. This leads to an increase in the serum lactate levels. Aerobic training reacts sensitively to this defect in the process of mitochondrial oxidative phosphorylation [11–13].

In practice, the LST is performed by measuring serum lactate levels before, during, and after absolute or relative exercise below the anaerobic threshold [14]. A significant increase in lactate is observed during exercise in MDs, while no such increase is observed in healthy individuals [15]. Various LST protocols with absolute and relative workloads have been proposed, which differ in particular with regard to test duration, the load intensity, and the various clinically-chemical parameters examined. The lack of standardization of these protocols is due to possible influencing factors, such as sex or the maximum strength of the subjects [11,13,16].

The LSTA involves performing exercise with a fixed absolute workload, for example a specified power output on a cycle ergometer. The goal of the LSTA is to determine whether an exercise-induced increases in lactate levels are a result of mitochondrial dysfunction [16].

On the other hand, the LSTR, involves performing exercise at a fixed relative workload, such as a percentage of the individual's maximal oxygen uptake (VO2max). The aim of the LSTR is to assess an individual's ability to perform in relation to their overall fitness level and to rule out the possibility that exercise-induced increases in lactate levels are due to factors other than mitochondrial dysfunction, such as reduced cardiovascular function or muscle weakness [16].

Although several studies have evaluated the diagnostic accuracy of the LSTA and LSTR, the use of different serum lactate cut-off's and exercise protocols in previous studies can make it difficult to interpret results from clinical practice. The diagnostic efficacy of these variations has not been thoroughly studied [17–21]. Due to the lack of detailed reviews evaluating the diagnostic accuracy of LSTs, particularly in relation to MDs, the aim of this study was to systematically review the available evidence in the literature and to perform a meta-analysis and assess the accuracy of LSTA and LSTR protocols for diagnosing MDs.

2. Materials and methods

The present systematic review and meta-analysis adhered to the guidelines established in the PRISMA-DTA (Preferred Reporting Items for a Systematic Review and Meta-Analysis of Diagnostic Test Accuracy Studies) statement [22]. Additionally, the study has been registered with PROSPERO under the registration number (CRD42022331710).

The objective of this review was based on the following research question:

What are the sensitivity (Sn), specificity (Sp), and diagnostic accuracy of the LSTA and LSTR in diagnosing MDs?

Table 1
PICOS criteria for inclusion in the systematic review and meta-analysis.

Acronym	Definition	Application of the criteria on the present study
P	Participants	Female and male patients with suspected mitochondrial disorders and received a clinical examination, muscle biopsy, biochemical, and genetic investigations as reference standards and received Lactate stress test used as index test
I	Intervention	The primary exposure in our inclusion criteria was the use of lactate stress test (index text) in the diagnosis of patients with suspected mitochondrial disorders
C	Comparator	Our analysis included any groups that were compared to our main population received lactate stress test (index test). Most frequently, these comparator groups were muscle biopsy, biochemical and genetic investigations (reference standards)
O	Outcomes	The outcomes of interest to answer the review objectives are following parameters: - Primary outcomes were sensitivity and specificity - We also collected or calculated data on PPV, NPV, LR+ and LR-
S	Study design	Study design of studies included were case-control and cohort diagnostic studies

Based on this objective, the PICOS [23] elements were defined, as summarized in Table 1.

2.1. Database search

Online databases (PubMed, Web of Science and Scopus) were reviewed from November 18th, 2021 until July 27th, 2022 for relevant studies. The search was restricted to human studies and those published in English. Full text articles not available, duplicate publications, and studies with missing data were excluded and the reason was described.

Our search strategy encompassed subject headings and free-text terms for the concepts "mitochondrial disease," "lactate stress test," and "diagnostic accuracy". The search strategy involved utilizing the fundamental Boolean search operators (AND, OR, and NOT) to construct the search query. Additionally, the search syntax has been customized to take advantage of the advanced search options available in each specific database. For detailed information, see the additional file Table S1 containing the search strategy.

2.2. Study inclusion/exclusion

We applied the following inclusion criteria to select studies: (1) the studies included had to involve human participants with mitochondrial diseases; (2) the participants should not have received any prior treatment; (3) the LST was used as the main diagnostic test, while muscle specimen or genetic testing served as the gold standard; (4) the study design had to be diagnostic studies; (5) studies had to provide sufficient original data for analysis.

Excluded were studies (1) involving subjects other than human beings or patients without MD; (2) where the participants did not undergo the LST; (3) involving participants who had previously been treated were excluded; (4) that have been published as reviews, case reports, letters and commentaries (5) not written in English and (6) that were republished.

2.3. Data extraction

Two separate researchers extracted the data independently. If disagreements occurred, a third researcher reviewed the data to reach consensus.

The pertinent information extracted from each study encompassed the following parameters: (1) First author's name; (2) The study's year of publication; (3) Country of origin; (4) Number of patients; (5) Mean age of patients; (6) Sex ratio of patients; (7) MD subtype (8) Reference standard test; (9) Index test classification according to the LST; (10) Timing and workload protocol; (11) Cut-off value for lactate measurement; (12) Diagnostic accuracy measures such as sensitivity (Sn), specificity (Sp), positive likelihood ratio (LR+), negative likelihood ratio (LR-), true positive (TP) and false positive (FP) rates, false negative (FN) and true negative (TN) rates, as well as positive predictive value (PPV) and negative predictive value (NPV); (13) test exercise equipment; (14) lactate testing sites; (15) lactate testing apparatus; (16) lactate measurement time; (17) the mean lactate level before the exercise; (18) the mean lactate level during exercise; (19) the mean lactate level during the post-exercise recovery period.

In case of multiple publications with overlapping sample data, those with the largest participant numbers or the most comprehensive information were preferred.

2.4. Process of assessing the methodological quality

We employed the QUADAS-2 tool (Quality Assessment Tool for Diagnostic Accuracy Studies-2) to evaluate the quality of data and the presence of bias in the included studies. This tool consists of four domains: patient selection, index test, reference standard, and flow and timing. Each domain was assessed for potential bias, categorized as low risk, high risk, or unclear risk. Additionally, the applicability of the first three domains was evaluated, graded as low risk, high risk, or unclear risk. The risk of bias and applicability for eligible studies were analyzed using Review Manager 5.4 software [24].

2.5. Statistical analysis

The study's heterogeneity was analyzed with Cochran Q and I^2 statistics. If the I^2 value was equal to or exceeded 50 %, the heterogeneity was deemed significant, and a random-effects model was deemed necessary. Conversely, if the I^2 value was below 50 %, indicating less heterogeneity, a fixed-effects model was utilized [25]. In order to create 2x2 tables, the available parameters were used to recalculate TP, FP, TN, and FN. Meta-DiSc 2.0 was utilized to apply the bivariate meta-analysis model, which determined the pooled Sn, Sp, LR+, LR-, and diagnostic odds ratio (DOR) [26,27].

Using the Sn and Sp data from each study at a single test threshold, the area under the curve (AUC) was calculated using the following method: First, the individual Sn and Sp values were plotted on a receiver operating characteristic (ROC) space. The AUC was then calculated by integrating the ROC curve, which plots the true-positive rate (Sn) against the false-positive rate (1-Sp). To generate the summary receiver operator characteristic (SROC) curve, we used the MIDAS command in Stata, which fits a bivariate random-effects model to account for study-level variability and provide a more accurate estimate of diagnostic accuracy [28].

Assessment of the diagnostic accuracy of LSTR and LSTA involved examining the Sn, Sp, and AUC values with corresponding 95 % confidence intervals (CI). Sensitivity analysis and subgroup analysis were conducted to explore potential factors contributing to heterogeneity. Threshold effects were assessed by analyzing the correlation coefficient between Sn and Sp in the bivariate model, with positive values indicating the presence of heterogeneity. Publication bias was assessed using Deeks' funnel plot, and a significance

level of $p < 0.05$ was employed.

3. Results

3.1. Study selection process

Initially, a total of 377 references were extracted. After reviewing the titles, 49 references were excluded due to data replication. After checking titles and abstracts, an additional 279 studies were excluded for being irrelevant, case reports, basic experiments, reviews, comments, non-English publications, or non-human studies.

Following a thorough review of the remaining 49 articles, 18 were excluded as they did not utilize the LST. Seven studies lacked assessments of diagnostic accuracy, four studies were excluded because the patients received treatment prior to the LST, and six studies were excluded due to overlapping participants.

Finally, 14 studies involving 1064 participants were deemed eligible for meta-analysis. Out of these, six studies included 793 participants undergoing the LSTA [15,29–33], and eight studies included 271 participants undergoing the LSTR [11,17,34–39] (Fig. 1).

3.2. Patient characteristics

The included studies were conducted between 1986 and 2022 and the number of participant ranged from 15 to 406 individuals. The included studies were conducted in Asia (Japan), Europe (Austria, Denmark, Germany, Sweden, and the United Kingdom), and the Americas (Canada and US).

The clinical subtypes included in our study were diverse and included Chronic Progressive External Ophthalmoplegia (CPEO), Kearns-Sayre Syndrome (KSS), CPEO-plus, Myoclonic Epilepsy with Ragged-Red Fibers (MERRF), Leber's Hereditary Optic

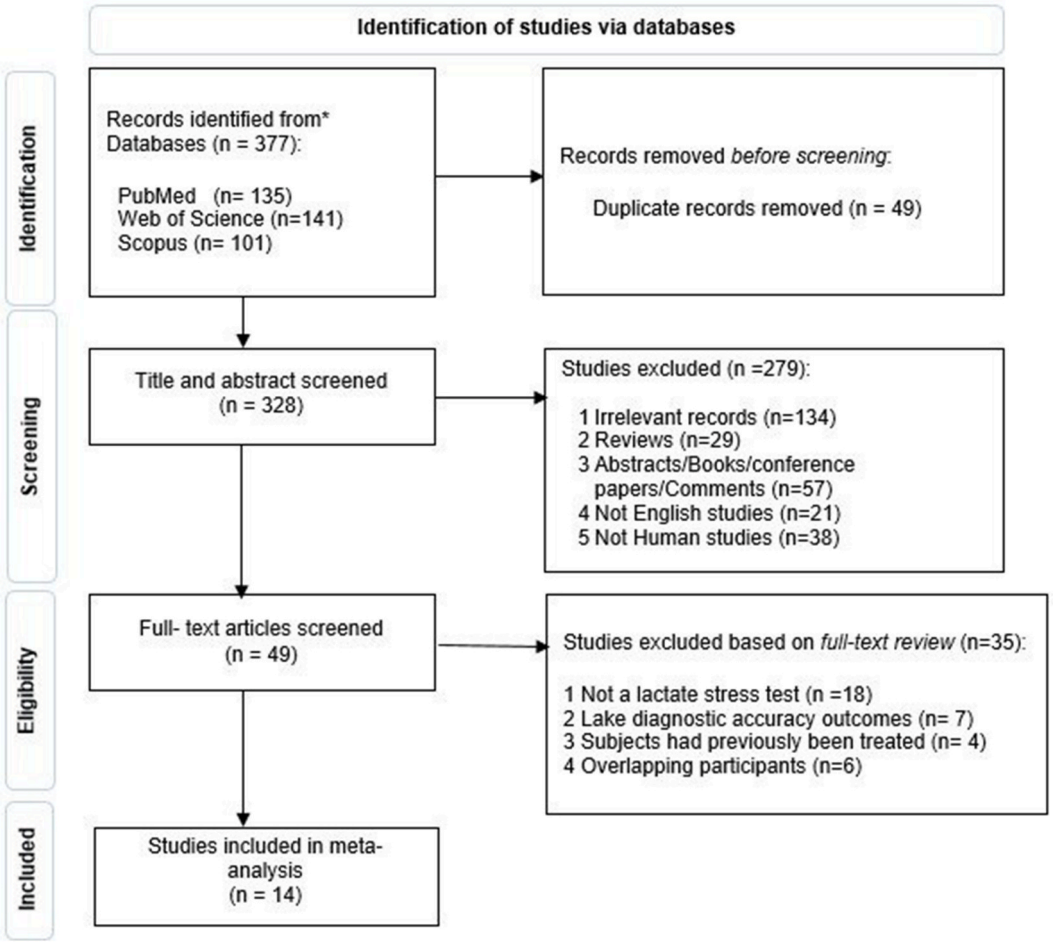


Fig. 1. Flow diagram of study selection process. n, number of studies.

Neuropathy (LHON), Progressive External Ophthalmoplegia (PEO), unspecified subtypes of mitochondriopathy (MCP), MELAS (Mitochondrial Encephalomyopathy, Lactic Acidosis, and Stroke-like episodes), Neuropathy, Ataxia, and Retinitis Pigmentosa (NARP), POLG-related disorders, and Leigh syndrome. More detailed information can be found in Table 2, which summarizes the basic characteristics of the included studies.

3.3. Lactate stress testing conditions

3.3.1. Test exercise equipment

Two studies used the MedGraphics CPE 2000 operated via a MedGraphics CPX/D [17,35]; 2 used the paddle-rate independent electronic bicycle ergometer E980 [15,31]; 1 used the Lode NV [11]; 1 used the Bosch ERG 551 [34]; 1 used the Douglas bags mass spectrometry [36]; 1 used the Monark 239E [38]; 1 used the Ergometer TX1 [32]; 1 used the StrengthErgo240 [33]; 1 used the handheld dynamometer [39] and in 3 studies the bicycle ergometer was not described [29,30,37].

3.3.2. Timing & workload protocol

Of six studies investigating LSTA, four studies used the same absolute power (W) and the same timing 30W for 15min [15,30–32]; 1 used 15W for 15min [33] and 1 used pedaling at a Rate of 10 km/h for 15min to produce 150/min heart rate [29].

Of eight studies investigating LSTR, five studies used the sub-anaerobic threshold exercise test (SATET) with several protocols, 2 reported using 90 % of the predicted workload for 15min [11,38]; 1 used 65 % VO2max for 20min [17]; 1 used for the first 2 loads 5 and 25W and for exhaustion 5, 10, 20 or 25W for 6–10min [34] and 1 used 5 or 10W every 1–2min for 10–15min to produce maximal heart rate (220-age) or exhaustion [35]. Three studies used the handgrip exercise test, 1 reported at 1/2 Hz for 3 min at 40 % of maximal voluntary contraction (MVC) force [37]; 1 used at 1/2 Hz for 6 min at 50 % of MVC force [39] and 1 used at the first 3min 30 % MVC and at the last 3min 50 % MVC [36].

3.3.3. Lactate testing sites

Of six studies investigating LSTA, five studies took blood from the medial cubital vein [16,30–33] and one study from the right radial artery of the forearm [29].

Of eight studies investigating LSTR, six studies took blood from the medial cubital vein [10,35–39] and two studies from the right radial artery of the forearm [11,34].

3.3.4. Lactate testing apparatus

One study did not describe the method of lactate measurement [38]. Five studies measured blood lactate using the automated enzymatic assay (Yellow Springs Instruments) [17,34–37]; two research papers with the Ektachrome Clinical Chemistry Slide [15,31]; one reported using the standard spectrophotometric determination [29]; one used the standard coupled-enzyme technique (Sigma L+ (lactate) 826-UV) [11]; one used commercial test kits (Boehringer Mannheim) [30]; one used the X20PRO Beckman Coulter [32]; one used the ABL90 FLEX blood gas analyzer [39] and 1 used the enzymatic method using lactate oxidase [33].

3.3.5. Lactate cut-off value

Of six studies investigating LSTA, two studies used the same peak lactate cut-off value, which was valued at $>$ 2.0 mM (18.0 mg/dL) [29,31]; 1 used $>$1.9 mM (17.1 mg/dL) [30]; one used $>$2.9 Mm (26.1 mg/dL) [32], one study used $>$2.13 mM (19.2 mg/dL) [33], two studies reported rest lactate cut-off value $>$ 1.4 mM (12.6 mg/dL) [30], and two $>$2.1 mM (18.9 mg/dL) [15].

Of the eight studies investigating LSTR, five studies reported rest lactate cut-off values, 2 reported $>$2.0 mM (18.0 mg/dL) [35,39]; 1 used $>$2.1 mM (18.9 mg/dL) [37]; one used $>$1.8 mM (16.2 mg/dL) [34]; one used $>$1.7 mM (15.2 mg/dL) [17], two studies reported peak lactate cut-off value $>$ 5.0 mM (45.0 mg/dL) [11] and two $>$2.9 mM (26.1 mg/dL) [38] and one study provided no description of lactate cut-off value [36].

3.3.6. Lactate rest

Of six studies using the LSTA, 3 studies measured resting blood lactate immediately prior to beginning the exercise [15,29,31] and 3 studies measured resting blood lactate in participants after 30min rest [30,32,33].

Of eight studies using the LSTR, five studies measured resting blood immediately before the test [11,34,36,38,39] and three studies provided no description of resting measurement time for participants prior to blood sampling [17,35,37].

3.3.7. Lactate during exercise

Six studies using the LSTA measured blood lactate at a 5min interval during exercise [15,29–33]. Of the eight studies that used the LSTR, two measured blood lactate at 2min intervals during exercise [34,39]; one study measured blood lactate immediately after the test [11]; one study measured blood lactate during the last 30 s of exercise at each workload [36]; one study measured blood lactate after 5min and 15min during the exercise test [38] and three studies provided no description of blood lactate measurement time [17, 35,37].

3.3.8. Lactate post exercise recovery

Of six studies using the LSTA, 3 measured blood lactate obtained after 15min recovery [15,31,32]; In one study, blood lactate levels were measured in the participants at 5, 15, 30, 45 and 75 min after completing the exercise [29] and two studies provided no

Table 2
Characteristics of the included studies and participants baseline demographics.

Study	Country	Number of participants	Sex ratio (f/m) P	Mean age	MD subtype	Reference standard	Test index	Timing & Workload Protocol	Lactate cut-off value	Sn (%)	Sp (%)	TP	FN	FP	TN	PPV	NPV	LR^+	LR^-
Petty RKH et al. (29)	United Kingdom	n(P) = 29 n(HC) = 14 n(DC)= NA	NA	NA	CPEO/KSS	Clinical manifestations, CK, EMG, Biochemical investigation of muscle biopsy	LSTA	Pedaling at a Rate of 10 Km/h for 15min to produce 150/min heart rate	Peak lactate >2.0 mM (18.0 mg/dL)	66	100	19	10	0	14	100	58	Inf	0.35
Nashef L and Lane RJM (11)	United Kingdom	n(P) = 6 n(HC) = 29 n(DC)= NA	3/3 15/14 NA	58 ± 12.76 NA NA	CPEO-plus	Clinical manifestations, CK, EMG, MRS, Histology investigations of muscle biopsy	LSTR	SATET (90 % of the predicted workload for 15min)	Peak lactate >5.0 mM (45.0 mg/dL)	100	93	6	0	2	27	75	100	14.5	0
Dandurand et al. (34)	Canada	n(P) = 8 n(HC) = 7 n(DC) = 4	4/4 3/4 1/3	44.5 ± 17.5 42.7 ± 12.1 37.3 ± 6.13	MERRF/ LHON/ CPEO/KSS	Clinical manifestations) Histology (RRF), Biochemical investigation of the muscle homogenate, Genetic testing	LSTR	SATET (First 2 loads: 5 and 25W Exhaustion: 5,10,20 or 25W 6–10min)	Rest lactate >1.8 mM (16.2 mg/dL)	25	57	2	6	3	4	40	40	0.58	1.31
Dengler R et al. (30)	Germany	n(P) = 20 n(HC) = 25 n(DC)= NA	5/15 16/9 NA	36.6 (15–57) 34.2 (21–75) NA	PEO	Clinical manifestations, Histology (RRF, AOE,AM, PI), Biochemical investigation of the muscle homogenate	LSTA	30W for 15min	Rest lactate >1.4 mM (12.6 mg/dL) Or Peak lactate >1.9 mM (17.1 mg/dL)	75	100	15	5	0	25	100	83	Inf	0.25
Taivassalo T et al. (36)	United States of America	n(P) = 13 n(HC) = 13 n(DC) = 11	7/6 6/7 5/6	37.5 ± 9.4 36.4 ± 9.2 34.2 ± 18.0	CPEO	Clinical manifestations, Biochemical investigation, Genetic testing	LSTR	Handgrip exercise test (The first 3min 30 % MVC, the last 3min 50 % MVC)	NA	77	100	10	3	0	13	100	81.3	Inf	0.23
Finsterer J and Milvay E. (15)	Austria	n(P) = 155 n(HC) = 62 n(DC) = 31	83/72 31/31 16/15	54.3 ± 16.0 42.6 ± 15.3 50.3 ± 15.2	MCP	Clinical manifestations, Histology (RRF), Biochemical investigation of the muscle	LSTA	30W for 15min	Rest lactate >2.1 mM (18.9 mg/dL)	66	94	103	52	4	58	98	36	10.3	0.36

(*continued on next page*)

Table 2 (*continued*)

Study	Country	Number of participants	Sex ratio (f/m) P	Mean age	MD subtype	Reference standard	Test index	Timing & Workload Protocol	Lactate cut-off value	Sn (%)	Sp (%)	TP	FN	FP	TN	PPV	NPV	LR^+	LR^-
						homogenate, Genetic testing													
Jensen T et al. (37)	Denmark	n(P) = 12 n(HC) = 12 n(DC) = 10	7/5 7/5 4/6	39 ± 5 39 ± 4 40 ± 3	CPEO	Clinical manifestations, CK, MRI, Histology (RRF, COX-negative), Genetic testing	LSTR	Handgrip exercise test (At 1/2 Hz for 3 min at 40 % of MVC force)	Rest lactate >2.1 mM (18.9 mg/dL)	50	58	6	6	5	7	54	54	1.20	0.86
Jeppesen TD et al. (17)	Denmark	n(P) = 15 n(HC) = 18 n(DC) = 10	8/7 8/10 3/7	38 ± 4 38 ± 3 36 ± 4	CPEO	Clinical manifestations, Histology (RRF), Genetic testing	LSTR	SATET (65 % VO_{2max} for 20min)	Rest lactate >1.7 mM (15.2 mg/dL)	27	94	4	11	1	17	80	61	4.80	0.78
Taivassalo T et al. (35)	United States of America	n(P) = 40 n(HC) = 32 n(DC)= NA	22/18 9/23 NA	37 ± 12 39 ± 8 NA	CPEO/ MELAS/ MERRF	Clinical manifestations, Biochemical investigation, Genetic testing	LSTR	SATET (5or10W every 1–2min for 10–15min to produce maximal heart rate (220-age) or exhaustion)	Rest lactate >2.0 mM (18.0 mg/dL)	35	100	14	26	0	32	100	55	Inf	0.65
Finsterer J and Milvay E. (31)	Austria	n(P) = 291 n(HC) = 115 n(DC) = 166	159/132 65/50 62/104	55.4 ± 15.7 43.1 ± 14.3 53.7 ± 16.2	MCP	Clinical manifestations, CK, EMG, Histology (RRF, AOE,AM, PI), Biochemical investigation of the muscle homogenate, Genetic testing	LSTA	30W for 15min	Peak lactate >2.0 mM (18.0 mg/dL)	66	84	192	99	18	97	91	49	4.22	0.40
Hammarén E et al. (38)	Sweden	n(P) = 9 n(HC) = 9 n(DC) = 10	8/1 8/2 7/2	49.33 ± 11.7 NA 50.60 ± 12.69	CPEO/ MELAS	Clinical manifestations, Histology (RRF), Genetic testing	LSTR	SATET (90 % of the predicted workload for 15min)	Peak lactate >2.9 mM (26.1 mg/dL)	78	100	7	2	0	9	100	82	Inf	0.22
Hanisch F et al. (32)	Germany	n(P) = 24 n(HC) = 37 n(DC) = 26	14/10 14/23 13/13	45 ± 14 45 ± 15 43 ± 15	CPEO/ MELAS	Histological examination, Biochemical investigation, Genetic testing	LSTA	30W for 15min	Peak lactate >2.9 Mm (26.1 mg/dL)	71	92	17	7	3	34	85	83	8.74	0.32

(*continued on next page*)

Table 2 (*continued*)

Study	Country	Number of participants	Sex ratio (f/m) P	Mean age	MD subtype	Reference standard	Test index	Timing & Workload Protocol	Lactate cut-off value	Sn (%)	Sp (%)	TP	FN	FP	TN	PPV	NPV	LR^+	LR^-
Løkken N et al. (39)	Denmark	n(P) = 29 n(HC) = 19 n(DC)= NA	15/14 8/11 NA	46 ± 13 47 ± 14 NA	CPEO/ NARP/ MERRF/ POLG/ KSS/ MELAS/ Leigh syndrom	Clinical manifestations, Genetic testing	LSTR	Handgrip exercise test (At 1/2 Hz for 6 min at 50 % of MVC force)	Rest lactate >2.0 mM (18.0 mg/dL)	41	95	12	17	1	18	92	51	7.86	0.62
Kurihara M et al. (33)	Japan	n(P) = 6 n(HC)= NA n(DC) = 15	0/6 NA 4/11	46.50 (37.0–51.5) NA 41.0 (28.5–55.5)	MERRF/ Leigh syndrome/ KSS/ NARP/ LHON	Clinical manifestations, Muscle histology Enzyme activity, Genetic testing	LSTA	15W for 15min	Peak lactate >2.13 mM (19.2 mg/dL) And L/P > 20 at peak lactate	50	93	3	3	1	14	75	82	7.50	0.54

P, patients with mitochondrial disorders; HC, Healthy controls; DC, Disease controls; CK, Creatine kinase; EMG, Electromyography; MRS, Magnetic resonance spectroscopy; RRF, ragged red muscle fibers; AOE, abnormal oxidative enzyme staining; AM, abnormal mitochondria; PI, paracristalline inclusions; SATET, the sub-anaerobic threshold exercise test; MVC, maximal voluntary contraction; Sn, sensitivity; Sp, specificity; TP, true positive; FN, false negative; FP, false positive; TN, true negative; PPV, Positive predictive value; NPV, negative predictive value; LR^+, positive likelihood ratio; LR^-, negative likelihood ratio; NA, not available; MCP: Mitochondriopathy; CPEO: Chronic Progressive External Ophthalmoplegia; KSS: Kearns-Sayre Syndrome; MERRF: Myoclonic epilepsy with ragged red fibers; MELAS: Mitochondrial Encephalomyopathy, Lactic Acidosis, and Stroke-like episodes; PEO: Progressive External Ophthalmoplegia; LHON: Leber Hereditary Optic Neuropathy; NARP: Neuropathy, Ataxia, and Retinitis Pigmentosa; POLG: Polymerase Gamma.

description of blood lactate measurement time during recovery.

Of eight studies using the LSTR, one measured blood at 30min after completion of exercise [11]; one study measured blood at 2, 4, 8, 16, 32, 48 and 64min following exercise [34]; one study measured blood at 5 and 10min post-exercise [36]; one study measured blood after 2, 10 and 20min of recovery phase [39] and four studies provided no description of lactate recovery blood measurement for participants [17,35,37,38].

Further details are provided in an additional file Table S2.

3.4. Study quality

All studies included had a low risk of bias. For patient selection, two studies were unclear, two studies had a high risk of bias, while ten studies didn't present any bias. Moreover, no risk of bias was detected in twelve and thirteen studies regarding the index test and reference standard as well as flow and timing respectively. Overall, most studies were considered to be of good quality, as less than 25 % were at high risk of bias in patient selection and index testing, as shown in the summary of risk of bias applicability concerns in Fig. 2 and concerns graph in Fig. 3.

3.5. Diagnostic accuracy of LSTA for MDs

The point estimates of the diagnostic accuracy of LSTA varied across six studies, with sensitivity ranging from 0.50 to 0.75 (I2 = 0 %) as shown in Fig. 4A and Sp ranging from 0.84 to 1.00 (I2 = 0 %) (Fig. 5A). These studies included a total of 793 participants, with 525 patients and 268 healthy controls.

A bivariate meta-analysis was conducted to summarize the findings. The summary estimates obtained were as follow: sensitivity of 0.67 (95 % confidence interval [CI] 0.62 to 0.72) and Sp of 0.93 (95 % CI 0.85 to 0.97). The bivariate I2 was 0 %, indicating no significant variability across the studies. The DOR was 26.63 (95 % CI 10.99 to 64.52), and the AUC was 0.70 (95 % CI 0.66 to 0.74), which suggests a reasonably discriminative ability.

The LR+ was 9.59 (95 % CI 4.27 to 21.53), indicating that a positive test result is almost ten times more likely in patients compared to healthy controls. The LR-was 0.36 (95 % CI 0.32 to 0.42), suggesting that a negative test result reduces the odds of the presence of the condition by 36 %.

There was no correlation between Sn and Sp in the bivariate model for this studies. Additionally, the scatter plots did not exhibit the

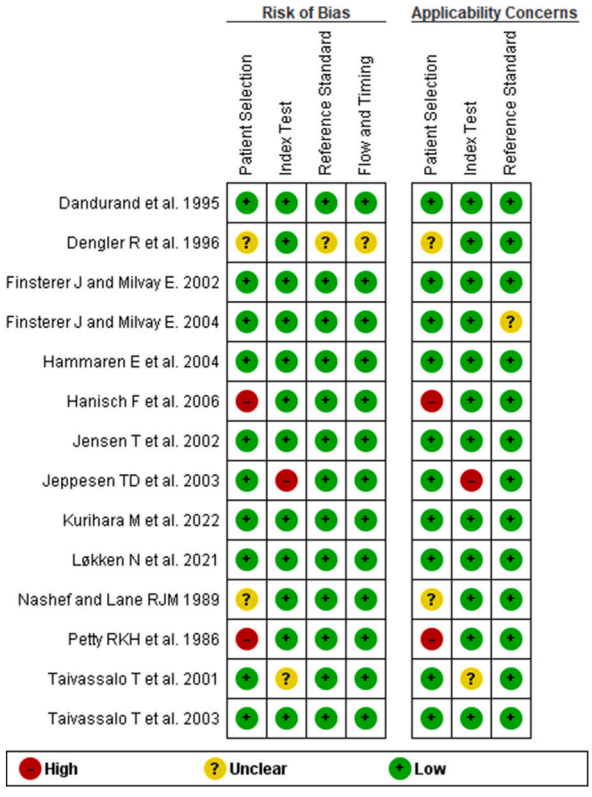


Fig. 2. Risk of bias and applicability concerns summary of included studies using the QUADAS-2 tool. QUADAS-2, Quality Assessment of Diagnostic Accuracy Studies.

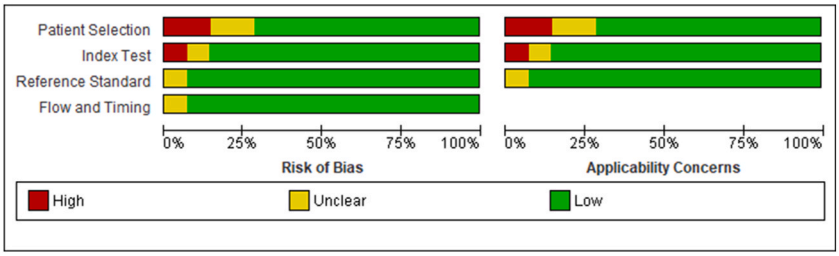


Fig. 3. Risk of bias and applicability concerns graph of included studies using the QUADAS-2 tool.

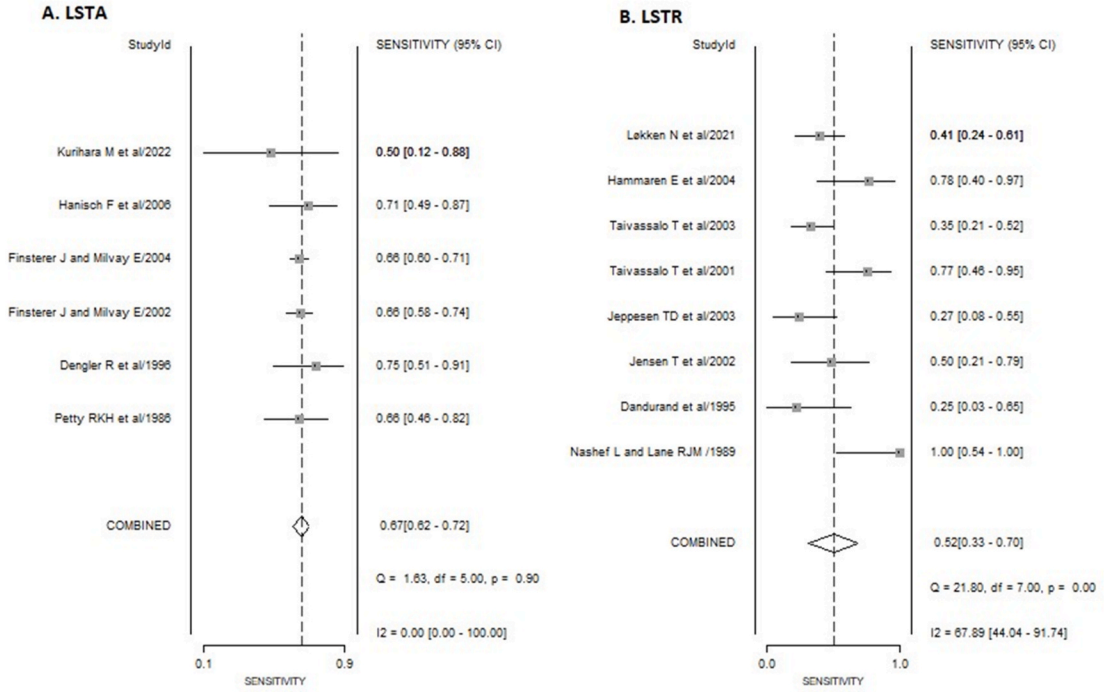


Fig. 4. Forest plot of sensitivity of individual studies for LSTA (A) and LSTR (B). TP, true positive; FP, false positive; FN, false negative; TN, true negative; 95 % CI, 95 % confidence interval.

characteristic "shoulder-arms" pattern in the SROC curve (as shown in Table 3 and Fig. 6A).

3.6. Diagnostic accuracy of LSTR for MDs

The point estimates for sensitivity of LSTR diagnostic accuracy varied from 0.25 to 1.00 (I2 = 45 %) across the eight studies as shown in Fig. 4B, while the point estimates for Sp ranged from 0.57 to 1.00 (I2 = 43 %) (Fig. 5B). These studies included a total of 271 participants, with 132 patients and 139 healthy controls.

The bivariate meta-analysis yielded summary estimates for Sn and Sp. The summary estimate for Se was 0.52 (95 % CI 0.33 to 0.70), and the summary estimate for Sp was 0.94 (95 % CI 0.79 to 0.99). The bivariate I2 was 48 %, indicating moderate heterogeneity. The DOR was 18.14 (95 % CI 2.99 to 109.85), and the AUC was 0.80 (95 % CI 0.76 to 0.83), which is close to 1.0. The LR+ was 9.28 (95 % CI 2.02 to 42.66), and the LR-was 0.52 (95 % CI 0.34 to 0.77). The correlation coefficient between Sn and Sp in the bivariate model was 0.35. The scatter plots in the SROC curve did not show a distinct "shoulder-arms" pattern (Table 3, Fig. 6B).

3.7. Diagnostic accuracy of overall LST for MDs

The range of point estimates for overall LST Sn across the fourteen studies was between 0.25 and 1.00, with an I2 value of 63 %. The range for Sp was between 0.57 and 1.00, with an I2 value of 29 %. The studies included a total of 1064 participants, consisting of 657 patients and 407 healthy controls.

Using bivariate meta-analysis, the pooled estimates for Sn of the LST were 0.59 (95 % CI 0.49 to 0.69), while the Sp estimates were

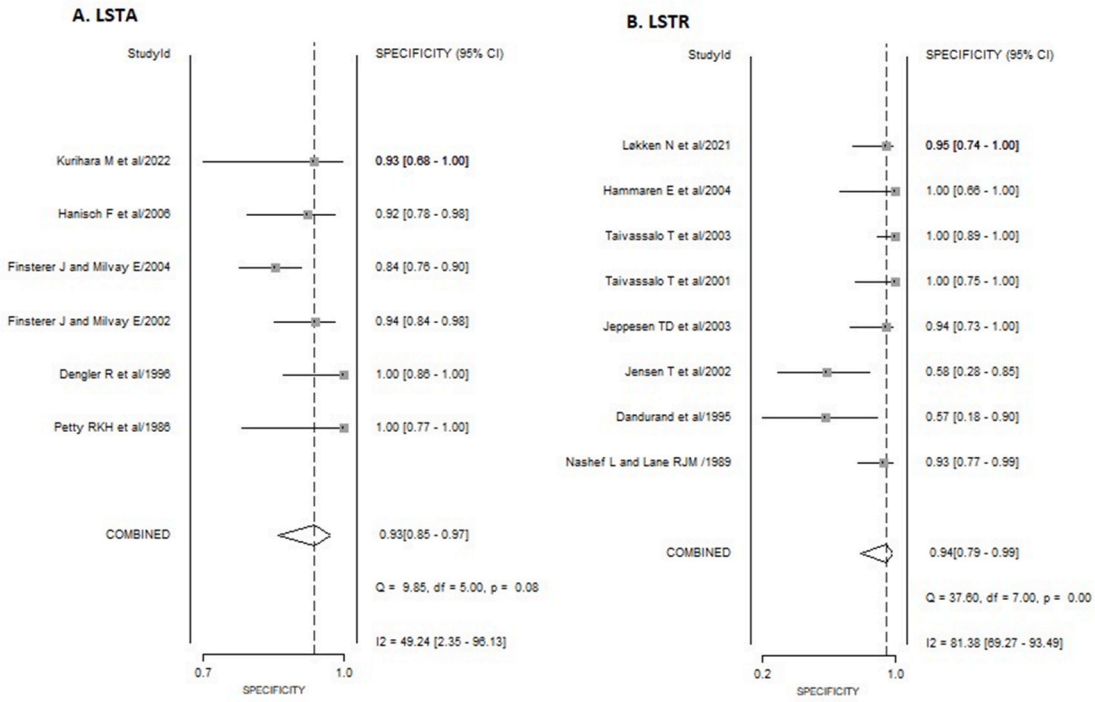


Fig. 5. Forest plot of specificity of individual studies for LSTA (A) and LSTR (B). TP, true positive; FP, false positive; FN, false negative; TN, true negative; 95 % CI, 95 % confidence interval.

Table 3
Summary of the pooled estimates (95 % CI) of LSTA, LSTR and overall LST in the diagnosis of MDs.

	Estimates of LSTA	Estimates of LSTR	Estimates of overall LST
***n* studies**	6	8	14
Number of subjects	793	271	1064
Sensitivity	0.67 (0.62, 0.72)	0.52 (0.33, 0.70)	0.59 (0.49–0.69)
Univariate I^2 sensitivity (%)	0	45	63
Specificity	0.93 (0.85, 0.97)	0.94 (0.79, 0.99)	0.94 (0.87–0.97)
Univariate I^2 specificity (%)	0	43	29
Bivariate I^2 (%)	0	48	47
Correlation	–	0.35	0.29
Positive likelihood ratio	9.59 (4.27, 21.53)	9.28 (2.02, 42.66)	9.82 (4.15–23.18)
Negative likelihood ratio	0.36 (0.32, 0.42)	0.52 (0.34, 0.77)	0.44 (0.34–0.56)
Diagnostic odds ratio	26.63 (10.99, 64.52)	18.14 (2.99, 109.85)	22.58 (8.11–62.93)
AUC	0.70 (0.66, 0.74)	0.80 (0.76, 0.83)	0.83 (0.80–0.86)

0.94 (95 % CI 0.87 to 0.97). The bivariate I2 value was 47 %, indicating moderate heterogeneity. The DOR was 22.58 (95 % CI 8.11 to 62.93), demonstrating a significant diagnostic accuracy. The AUC was 0.83 (95 % CI 0.80 to 0.86), indicating a good overall performance of the LST. The positive likelihood ratio (LR+) was 9.82 (95 % CI 4.15 to 23.18), and the LR-was 0.44 (95 % CI 0.34 to 0.56). The correlation coefficient between Sn and Sp in the bivariate model was 0.29. The scatter plots did not exhibit a "shoulder-arms" pattern in the SROC curve (Table 3, Fig. 7).

3.8. Subgroup analysis and metaregression

To assess heterogeneity and the influence of individual studies on the overall results, we conducted a step-by-step exclusion of these studies. Additionally, we conducted exploratory subgroup analyses to evaluate the influence of various factors, including the protocols of LST, cut-off value, sample size, gender, and age. Our findings indicated that the cut-off value could potentially be the primary source of heterogeneity within the LSTR group (Table 4).

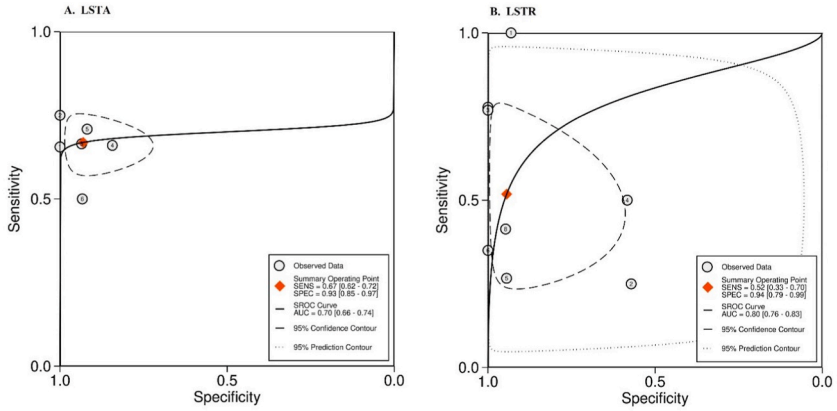


Fig. 6. Sensitivity, specificity, and summary receiver operating characteristic (SROC) curves LSTA (A) and LSTR (B) in the diagnosis of MDs. **A, LSTA**. 1: study by Petty RKH et al.; 2: study by Dengler R et al.; 3: study by Finsterer J and Milvay E. (2002); 4: study by Finsterer J and Milvay E. (2004); 5: study by Hanisch F et al.; 6: study by Kurihara M et al. **B, LSTR.** 1: study by Nashef L and Lane RJM; 2: study by Dandurand et al.; 3: study by Taivassalo T et al. (2001); 4: study by Jensen T et al.; 5: study by Jeppesen TD et al.; 6: study by Taivassalo T et al. (2003); 7: study by Hammarén E et al.; 8: study by Løkken N et al. AUC, area under the curve.

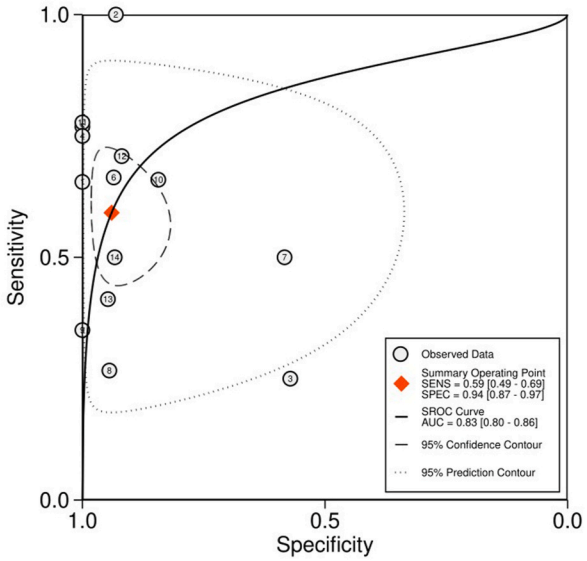


Fig. 7. Sensitivity, specificity and summary receiver operating characteristic (SROC) curves of overall LST for MDs 1: study by Petty RKH et al.; 2: study by Nashef L and Lane RJM; 3: study by Dandurand et al.; 4: study by Dengler R et al.; 5: study by Taivassalo T et al. (2001); 6: study by Finsterer J and Milvay E. (2002); 7: study by Jensen T et al.; 8: study by Jeppesen TD et al.; 9: study by Taivassalo T et al. (2003); 10: study by Finsterer J and Milvay E. (2004); 11: study by Hammarén E et al.; 12: study by Hanisch F et al.; 13: study by Løkken N et al.; 14: study by Kurihara M et al.

3.9. Publication bias

Deeks' funnel plot asymmetry test was utilized to assess bias in the study. The vertical axis of the funnel plot represents the inverse of the square root of the effective sample size ($1/\sqrt{ESS}$), while the horizontal axis represents the DOR. The overall results of the analysis, as depicted in Fig. 8, indicated that the Deeks' funnel plot exhibited symmetry with a p-value of 0.68. This finding suggests that minimal publication bias exist in the meta-analysis.

Similarly, for the LSTA and LSTR Deeks' funnel plots, shown in Figs. 9 and 10 respectively, the corresponding p-values were 0.17 and 0.44, respectively. These results indicate a symmetrical funnel shape, suggesting the absence of publication bias in these specific analyses.

Table 4
Subgroup analyses according to the type of protocols of LST, cut-off value, sample size, gender, and age.

Subgroup	LSTA							LSTR						
		n studies	Sensitivity	p-value	Specificity	p-value	Global test comparison		n studies	Sensitivity	p-value	Specificity	p-value	Global test comparison
Protocols of LST	30 W for 15min	4	0.67 (0.63–0.71)	0.64	0.92 (0.83–0.96)	0.33	0.56	SATET	5	0.48 (0.26–0.71)	0.69	0.95 (0.77–0.99)	0.67	0.81
	Other	2	0.63 (0.46–0.77)		0.97 (0.78–0.99)			Handgrip exercise test	3	0.55 (0.29–0.79)		0.92 (0.56–0.99)		
Cut-off value	Rest lactate	2	0.67 (0.61–0.74)	0.75	0.95 (0.87–0.98)	0.15	0.35	Rest lactate	5	0.37 (0.28–0.48)	<0.05	0.9 (0.67–0.97)	0.38	<0.05
	Peak lactate	4	0.66 (0.61–0.72)		0.88 (0.75–0.95)			Peak lactate	2	0.87 (0.59–0.97)		0.97 (0.69–0.99)		
Participants (>150)	Yes	2	0.66 (0.62–0.71)	0.71	0.88 (0.79–0.94)	0.09	0.22	Yes	–	–	–	–	–	–
	No	4	0.68 (0.57–0.77)		0.96 (0.88–0.98)			No	–	–		–		
Age matched	Yes	2	0.68 (0.54–0.79)	0.82	0.95 (0.82–0.98)	0.55	0.82	Yes	4	0.47 (0.29–0.66)	0.66	0.95 (0.53–0.99)	0,68	0.74
	No	4	0.66 (0.62–0.71)		0.92 (0.82–0.96)			No	3	0.41 (0.21–0.64)		0.98 (0.45–1.00)		
Gender matched	Yes	2	0.68 (0.54–0.79)	0.82	0.95 (0.82–0.98)	0.55	0.82	Yes	5	0.52 (0.29–0.73)	0,95	0.89 (0.64–0.97)	0,28	0.55
	No	4	0.66 (0.62–0.71)		0.92 (0.82–0.96)			No	3	0.53 (0.25–0.79)		0.97 (0.84–0.99)		

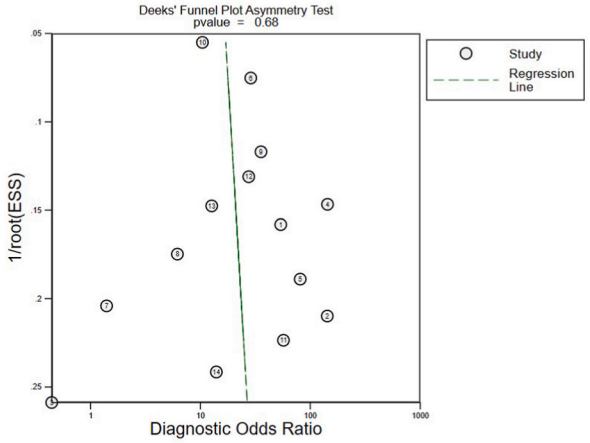


Fig. 8. Overall LST Deeks' funnel plot for detecting publication bias. 1: study by Petty RKH et al.; 2: study by Nashef L and Lane RJM; 3: study by Dandurand et al.; 4: study by Dengler R et al.; 5: study by Taivassalo T et al. (2001); 6: study by Finsterer J and Milvay E. (2002); 7: study by Jensen T et al.; 8: study by Jeppesen TD et al.; 9: study by Taivassalo T et al. (2003); 10: study by Finsterer J and Milvay E. (2004); 11: study by Hammarén E et al.; 12: study by Hanisch F et al.; 13: study by Løkken N et al.; 14: study by Kurihara M et al.

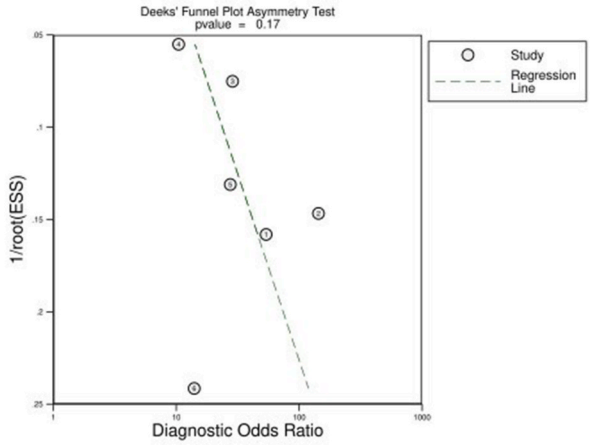


Fig. 9. LSTA Deeks' funnel plot for detecting publication bias. 1: study by Petty RKH et al.; 2: study by Dengler R et al.; 3: study by Finsterer J and Milvay E. (2002); 4: study by Finsterer J and Milvay E. (2004); 5: study by Hanisch F et al.; 6: study by Kurihara M et al.

4. Discussion

To the best of our knowledge, this is the first systematic review with meta-analysis that collects the existing evidence on the accuracy parameters of LSTA and LSTR. LST is a widely accepted method for first-line MD screening. However, there is an ongoing debate surrounding the methods of LST handling, including the choice of the adequate workload (LSTR or LSTA), the intensity of the physical effort, the duration of the workload, the optimal blood sampling site and timing for lactate measurement [40,41].

This meta-analysis has demonstrated that both LSTA and LSTR tests are valid tools for MDs diagnosis. The analysis showed that both LSTA and LSTR have nearly the same Sp. However, LSTA is more sensitive than LSTR, while LSTR has higher AUC than LSTA, suggesting that it may have better overall diagnostic performance.

The LSTA was introduced in the 1980s and has been compared to other methods such as the LSTR [42]. A comprehensive analysis of prior research examining the diagnostic efficacy of LSTA performed on cycle ergometers for the detection of MDs revealed non-significant heterogeneity (I2 = 0 %) in terms of study participants, exercise protocols, and threshold values [16,29–33].

Our analysis of six eligible studies for meta-analysis [16,29–33] showed a pooled Sn of 67 % and a pooled Sp of 93 % for the LSTA. The pooled AUC was 0.70. The validity, reliability, and reproducibility of the LSTA are consistent with previous findings. In this context, Finsterer [43] reported a reproducibility of 65 % for the LSTA in patients with MDs and 78 % in disease control subjects.

On the other hand, the LSTR enables a more individualized and focused assessment of mitochondrial activity [16]. The diagnostic accuracy of LSTR in the eight studies [11,17,34–39] showed a pooled Sn of 0.52 (95 % CI: 0.33, 0.70) and a specificity of 0.94 (95 % CI: 0.79, 0.99). The LR+ 9.28 (95 % CI: 2.02, 42.66), indicating a strong diagnostic utility.

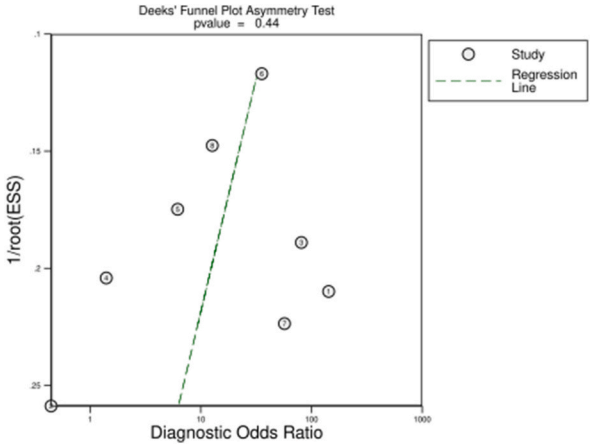


Fig. 10. LSTR Deeks' funnel plot for detecting publication bias. 1: study by Nashef L and Lane RJM; 2: study by Dandurand et al.; 3: study by Taivassalo T et al. (2001); 4: study by Jensen T et al.; 5: study by Jeppesen TD et al.; 6: study by Taivassalo T et al. (2003); 7: study by Hammarén E et al.; 8: study by Løkken N et al. AUC, area under the curve.

Interestingly, there was a moderate positive correlation (r = 0.35) between LSTR Sn and disease severity, suggesting that LSTR may be more effective in diagnosing advanced cases of MDs. However, the heterogeneity between the included studies was high, as indicated by the I2 values of 45 % and 43 % for Sn and Sp, respectively, and the bivariate I2 of 48 %. Based on the subgroup analysis and metaregression, we considered that cut-off values could be the main cause of heterogeneity for the LSTR.

Significantly, there is a notable variation in the cut-off values of lactate employed to predict MDs across the eight studies, with differences observed for both resting lactate values (ranging from 1.7 mmol/L to 2.1 mmol/L) and peak lactate values (ranging from 2.9 mmol/L to 5.0 mmol/L). This variation can largely be attributed to the utilization of different analyzers, including the standard coupled-enzyme technique, automated enzymatic assay, and ABL90 FLEX blood gas analyzer.

Taking into account the prevalence of MDs estimated around 1 in 5000 and the analysis of this previous results reveals some notable findings. The LSTA demonstrates a slightly higher Sn (67 %) compared to the LSTR (52 %). This implies that the LSTA is more effective in detecting true positives, although there remains a risk of false negatives. On the other hand, the LSTR exhibits a higher Sp (94 %) compared to the LSTA (93 %). This indicates that the LSTR better excludes false positives. Additionally, the LSTR had a higher AUC value (0.80) than the LSTA (0.70), suggesting an overall improved capability to distinguish between positive and negative cases. In conclusion, the LSTR compared to LSTA appears to have a slightly higher diagnostic performance, characterized by slightly lower Sn but higher Sp, and higher AUC.

In most cases, researchers prefer to perform LSTR due to its reliance on an individual's maximal aerobic capacity (VO2max). This preference is reflected in the higher number of eligible studies obtained in this work compared to those utilizing LSTA. This means that the workload is adjusted according to each individual's aerobic capacity, which allows for better control of exercise intensity and facilitates the comparison of results between different patients.

These results suggest moderate diagnostic performance for the overall LST. Furthermore, it has advantages over muscle biopsies, which is an invasive and frequently challenging procedure [39].

Finally, the comparison of workload and timing protocols between eight studies investigating LSTR showed that five studies used the SATET [11,17,34,35,38] with several protocols each one and for six studies investigating LSTA, four studies used the same absolute (W) power 30W and the same timing 15min [16,30–33].

The optimal blood sampling site for lactate testing is the medial cubital vein compared to the right radial artery of the forearm for both tests LSTA and LSTR and the results shows that for five out six studies investigating LSTA [16,30–33] and six out eight studies investigating LSTR [10,35–39] used the medial cubital vein.

The results showed a significant increase in lactate levels in MD patients during both LSTA and LSTR. In addition, compared to control subjects, lactate recovery was delayed in these patients. It is noteworthy to highlight that results from several studies varied significantly due to various exercise protocols, lactate thresholds, and measuring approaches. Hence, a consistent methodology should be used to increase the precision of lactate testing in assessing MDs during exercise.

Given the results, it is challenging to conclude with certainty which test is more adequate because both the LSTA and the LSTR have benefits and drawbacks. The choice of test may be influenced by equipment accessibility and the patient's unique aerobic capability, among others.

In summary, the following recommendations should be considered for standardization of LST procedures.

1. LST should be handled by a physician or other qualified professional.
2. The process should be explained and the patient familiar with tools and task.
3. Blood at rest should be drawn using a lancet or other blood sample tool, preferably, from the medial cubital vein.
4. If using the LSTR, the effort should be modified according to the patient's unique aerobic capacity.

5. During exercise, blood lactate should be repeatedly measured. Use a 2-min interval for LSTR and a 5-min period for LSTA between measurements.
6. At least 15 min should be allocated for a recovery interval so that blood lactate levels can be checked once more.
8. All unusual events should be recorded during the test, including the patient's symptoms, and blood pressure.
9. When employing LSTR, the patient's individual lactate threshold and other pertinent variables should be identified. However, lactate cut-off value for LSTA at rest can be (2.1 mM) and at peak (2.9 mM).

It is crucial to mention that this is only a broad protocol concluded on the basis of results obtained and may need to be modified depending on the demands of each patient and testing apparatus.

The strengths of the presented review and meta-analysis lie in the exhaustive literature search strategy, ensuring that a comprehensive range of relevant studies was captured and the double review process, whereby multiple reviewers independently assessed and screened the selected studies. This rigorous approach helps to reduce bias and enhances the reliability of the included studies. Furthermore, we conducted a meta-regression analysis, which allowed us to explore potential sources of heterogeneity among the included studies. These strengths contribute to the reliability, validity, and relevance of our findings, thereby enhancing the overall quality of the study.

It is important to also acknowledge the limitations of this study, it is challenging to find a perfectly matched control group of healthy individuals for validation purposes, because they may differ in terms of physical endurance capacity and lifestyle and 5/14 studies used healthy controls, while 9/14 studies included patients with non-mitochondrial conditions. This variation can lead to differences in specificity and potentially increase it when healthy controls are used. In clinical practice, the LST is used in patients with symptoms suggestive of MD. The fact that our review is based on studies with healthy controls may not accurately reflect real-world specificity. Future studies should focus on more clinically relevant control groups to better assess the diagnostic accuracy of the LST in practice.

Another potential another limitation relates to the overrepresentation of certain MD subtypes, particularly Chronic Progressive External Ophthalmoplegia (CPEO) and Kearns-Sayre Syndrome (KSS). Of the 14 studies included, a significant number predominantly involved patients with CPEO/KSS or related subtypes.

This disproportionate representation could bias our results and limit the generalizability of our findings across all MD subtypes. The high prevalence of CPEO/KSS in the included studies could bias the diagnostic accuracy metrics of the LST towards these specific subtypes. Consequently, our results may not be fully transferable to other less prevalent subtypes, such as MELAS, NARP, LHON) and Leigh Syndrome.

To address this limitation, we propose to perform subgroup analyses if possible, to evaluate the diagnostic accuracy of LST specifically for other MD subtypes. This approach will help to delineate whether the LST sensitivity/specificity varies significantly for different subtypes.

It is suggested that future studies should aim for a more balanced inclusion of various MD subtypes to improve the generalizability of the results. Further studies should investigate the diagnostic accuracy of the LST in less represented subtypes to allow a comprehensive assessment.

5. Conclusions

This is the first systematic review and meta-analysis evaluating accuracy parameters of the LSTR and LSTA for the detection of MDs. This meta-analysis demonstrated that both LSTA and LSTR tests are valid tools for MD diagnosis and have nearly the same Sp. However, LSTA is more sensitive than LSTR, while LSTR has a higher AUC than LSTA, suggesting that it may have better overall diagnostic performance. Generally, researchers prefer to perform LSTR due to its reliance on an individual's maximal aerobic capacity (VO2max). This preference is reflected in the higher number of eligible studies included in this analysis.

Further research is needed to assess whether LSTR is effective for the overall diagnosis of MDs than the LSTA. Until then, it is also possible to apply the LSTA, This method can be used when the patient's maximal aerobic capacity is not known or cannot be reliably evaluated. Large-scale studies are needed to optimize cut-off points and to harmonize protocols for both tests.

CRediT authorship contribution statement

Sara El Guessabi: Writing – review & editing, Writing – original draft, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Redouane Abouqal:** Validation, Supervision, Software, Methodology, Formal analysis. **Azeddine Ibrahimi:** Validation. **Ghizlane Zouiri:** Validation. **Fatima Sfifou:** Validation. **Josef Finsterer:** Writing – review & editing, Validation, Supervision. **Yamna Kriouile:** Validation, Supervision.

Patient consent for publication

Not applicable.

Data availability

All data generated or analyzed during this study are included in this published article and its supplementary materiels files.

Ethics approval

No ethical approval was be needed because data from previous published studies in which informed consent was obtained by primary investigators were be retrieved and analyzed.

Funding

No funding was received for conducting this study.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

I thank Professor ABOUQAL and Professor FINSTERER for their expertise and assistance throughout all aspects of our study and for their help in writing the manuscript.

Abbreviations

95 % CI	95 % Confidence Interval
AM	Abnormal mitochondria
AOE	Abnormal oxidative enzyme staining
AUC	Area under the ROC Curve
CK	Creatine Kinase
CPEO	Chronic Progressive External Ophthalmoplegia
DC	Disease controls
DNA	Deoxyribonucleic Acid
DOR	Diagnostic Odds Ratio
EMG	Electromyography
F	Female
FN	False Negative
FP	False Positive
HC	Healthy controls
KSS	Kearns-Sayre Syndrome
LHON	Leber Hereditary Optic Neuropathy
LR-:	Negative Likelihood Ratio
LR+	Positive Likelihood Ratio
LST	Lactate Stress Test
LSTA	Lactate Stress Test under Absolute workload
LSTR	Lactate Stress Test under relative workload
M	Male
MCP	Mitochondriopathy
MDs	Mitochondrial Disorders
MELAS	Mitochondrial Encephalomyopathy, Lactic Acidosis, and Stroke-like episodes
MERRF	Myoclonic Epilepsy with Ragged Red Fibers
MRS	Magnetic resonance spectroscopy
mtDNA	Mitochondrial DNA
MVC	Maximal Voluntary Contraction
NARP	Neuropathy, Ataxia, and Retinitis Pigmentosa
nDNA	Nuclear DNA
NGS	Next-Generation Sequencing
NPV	Negative Predictive Value
P	Patients with MDs
PI	Paracristalline inclusions
PICOS	Population, Intervention, Comparison, Outcomes and Study
POLG	Polymerase Gamma
PPV	Positive Predictive Value
PRISMA-DTA	Preferred Reporting Items for a Systematic Review and Meta-Analysis of Diagnostic Test Accuracy Studies
PROSPERO	International Prospective Register of Systematic Reviews

QUADAS-2 Quality Assessment of Diagnostic Accuracy Studies-2
RRF Ragged red fibers
SATET sub-anaerobic threshold exercise test
Sn Sensitivity
Sp: Specificity
SROC: Summary Receiver Operator Characteristic
TN True Negative
TP True Positive

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e39648.

References

[1] Y.S. Ng, D.M. Turnbull, Mitochondrial disease: genetics and management, J. Neurol. 263 (2016) 179–191, https://doi.org/10.1007/s00415-015-7884-3.
[2] S. Kanungo, J. Morton, M. Neelakantan, K. Ching, J. Saeedian, A. Goldstein, Mitochondrial disorders, Ann. Transl. Med. 6 (2018) 475, https://doi.org/10.21037/atm.2018.12.13.
[3] M.K. Koenig, Presentation and diagnosis of mitochondrial disorders in children, Pediatr. Neurol. 38 (2008) 305–313, https://doi.org/10.1016/j.pediatrneurol.2007.12.001.
[4] S.T. Ahmed, L. Craven, O.M. Russell, D.M. Turnbull, A.E. Vincent, Diagnosis and treatment of mitochondrial myopathies, Neurother J Am Soc Exp Neurother 15 (2018) 943–953, https://doi.org/10.1007/s13311-018-00674-4.
[5] S. Parikh, A. Goldstein, M.K. Koenig, F. Scaglia, G.M. Enns, R. Saneto, et al., Diagnosis and management of mitochondrial disease: a consensus statement from the Mitochondrial Medicine Society, Genet. Med. 17 (2015) 689–701, https://doi.org/10.1038/gim.2014.177.
[6] J. Finsterer, S. Shorny, J. Capek, C. Cerny-Zacharias, B. Pelzl, R. Messner, et al., Lactate stress test in the diagnosis of mitochondrial myopathy, J. Neurol. Sci. 159 (1998) 176–180, https://doi.org/10.1016/s0022-510x(98)00170-1.
[7] M. Schmidt, M. Kunkel, P. Schuff-Werner, M. Naumann, H. Reichmann, C.D. Reimers, Standardisierte aerobe Gehbelastung auf dem Laufband bei Gesunden sowie Patienten mit mitochondrialen und nichtmitochondrialen Myopathien, Nervenarzt 68 (1997) 831–835, https://doi.org/10.1007/s001150050201.
[8] R.J.M. Lane, M.C. Barrett, D. Woodrow, J. Moss, R. Fletcher, L.C. Archard, Muscle fibre characteristics and lactate responses to exercise in chronic fatigue syndrome, J. Neurol. Neurosurg. Psychiatry 64 (1998) 362–367, https://doi.org/10.1136/jnnp.64.3.362.
[9] A. Chan, H. Reichmann, A. Kögel, A. Beck, R. Gold, Metabolic changes in patients with mitochondrial myopathies and effects of coenzyme Q10 therapy, J. Neurol. 245 (1998) 681–685, https://doi.org/10.1007/s004150050267.
[10] T.D. Jeppesen, M. Schwartz, D.B. Olsen, F. Wibrand, T. Krag, M. Dunø, et al., Aerobic training is safe and improves exercise capacity in patients with mitochondrial myopathy, Brain J Neurol 129 (2006) 3402–3412, https://doi.org/10.1093/brain/awl149.
[11] L. Nashef, R.J. Lane, Screening for mitochondrial cytopathies: the sub-anaerobic threshold exercise test (SATET), J. Neurol. Neurosurg. Psychiatry 52 (1989) 1090–1094, https://doi.org/10.1136/jnnp.52.9.1090.
[12] A.H.V. Schapira, Mitochondrial disorders, Curr. Opin. Neurol. 10 (1997) 43.
[13] S. Zierz, S. Meessen, F. Jerusalem, Lactat- und Pyruvatblutspiegel in der Diagnostik mitochondrialer Myopathien, Lact- Pyruvatblutspiegel Diagn Mitochondrialer Myopathien 60 (1989) 545–548.
[14] J. Finsterer, S. Zarrouk-Mahjoub, Biomarkers for detecting mitochondrial disorders, J. Clin. Med. 7 (2018) 16, https://doi.org/10.3390/jcm7020016.
[15] J. Finsterer, E. Milvay, Lactate stress testing in 155 patients with mitochondriopathy, Can J Neurol Sci J Can Sci Neurol 29 (2002) 49–53, https://doi.org/10.1017/s0317167100001712.
[16] J. Finsterer, E. Milvay, Diagnostic yield of the lactate stress test in respiratory chain disorders under absolute and relative workload, J. Neurosci. Methods 108 (2001) 65–70, https://doi.org/10.1016/s0165-0270(01)00371-5.
[17] T.D. Jeppesen, D. Olsen, J. Vissing, Cycle ergometry is not a sensitive diagnostic test for mitochondrial myopathy, J. Neurol. 250 (2003) 293–299, https://doi.org/10.1007/s00415-003-0993-4.
[18] Improvement of abnormal pyruvate metabolism and cardiac conduction defect with coenzyme Q10 in Kearns-Sayre syndrome - PubMed n.d. https://pubmed.ncbi.nlm.nih.gov/3974895/. (Accessed 13 February 2023).
[19] S. Ogasahara, Y. Nishikawa, S. Yorifuji, F. Soga, Y. Nakamura, M. Takahashi, et al., Treatment of Kearns-Sayre syndrome with coenzyme Q10, Neurology 36 (1986) 45–53, https://doi.org/10.1212/wnl.36.1.45.
[20] Y. Nishikawa, M. Takahashi, S. Yorifuji, Y. Nakamura, S. Ueno, S. Tarui, et al., Long-term coenzyme Q10 therapy for a mitochondrial encephalomyopathy with cytochrome c oxidase deficiency: a 31P NMR study, Neurology 39 (1989) 399–403, https://doi.org/10.1212/wnl.39.3.399.
[21] K. Abe, Y. Matsuo, J. Kadekawa, S. Inoue, T. Yanagihara, Measurement of tissue oxygen consumption in patients with mitochondrial myopathy by noninvasive tissue oximetry, Neurology 49 (1997) 837–841, https://doi.org/10.1212/wnl.49.3.837.
[22] the PRISMA-DTA Group, M.D.F. McInnes, D. Moher, B.D. Thombs, T.A. McGrath, P.M. Bossuyt, et al., Preferred reporting Items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement, JAMA 319 (2018) 388–396, https://doi.org/10.1001/jama.2017.19163.
[23] C.M. da Costa Santos, C.A. de Mattos Pimenta, M.R.C. Nobre, The PICO strategy for the research question construction and evidence search, Rev Lat Am Enfermagem 15 (2007) 508–511, https://doi.org/10.1590/s0104-11692007000300023.
[24] QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies - PubMed n.d. https://pubmed.ncbi.nlm.nih.gov/22007046/. (Accessed 20 June 2023).
[25] J. Dinnes, J. Deeks, J. Kirby, P. Roderick, A methodological review of how heterogeneity has been examined in systematic reviews of diagnostic test accuracy, Health Technol Assess Winch Engl 9 (2005) 1–113, https://doi.org/10.3310/hta9120, iii.
[26] J.B. Reitsma, A.S. Glas, A.W.S. Rutjes, R.J.P.M. Scholten, P.M. Bossuyt, A.H. Zwinderman, Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews, J. Clin. Epidemiol. 58 (2005) 982–990, https://doi.org/10.1016/j.jclinepi.2005.02.022.
[27] M.N. Plana, I. Arevalo-Rodriguez, S. Fernández-García, J. Soto, M. Fabregate, T. Pérez, et al., Meta-DiSc 2.0: a web application for meta-analysis of diagnostic test accuracy data, BMC Med. Res. Methodol. 22 (2022) 306, https://doi.org/10.1186/s12874-022-01788-2.
[28] B. Dwamena, MIDAS: Stata module for meta-analytical integration of diagnostic test accuracy studies, Stat Softw Compon (2009).
[29] R.K. Petty, A.E. Harding, J.A. Morgan-Hughes, The clinical features of mitochondrial myopathy, Brain J Neurol 109 (Pt 5) (1986) 915–938, https://doi.org/10.1093/brain/109.5.915.
[30] R. Dengler, K. Wohlfarth, S. Zierz, M. Jobges, M. Schubert, Muscle fatigue, lactate, and pyruvate in mitochondrial myopathy with progressive external ophthalmoplegia, Muscle Nerve 19 (1996) 456–462, https://doi.org/10.1002/(SICI)1097-4598(199604)19:4<456::AID-MUS5>3.0.CO;2-B.

[31] J. Finsterer, E. Milvay, Stress lactate in mitochondrial myopathy under constant, unadjusted workload, Eur. J. Neurol. 11 (2004) 811–816, https://doi.org/10.1111/j.1468-1331.2004.00859.x.
[32] F. Hanisch, T. Müller, A. Muser, M. Deschauer, S. Zierz, Lactate increase and oxygen desaturation in mitochondrial disorders–evaluation of two diagnostic screening protocols, J. Neurol. 253 (2006) 417–423, https://doi.org/10.1007/s00415-006-0987-0.
[33] M. Kurihara, Y. Sugiyama, M. Tanaka, K. Sato, A. Mitsutake, H. Ishiura, et al., Diagnostic values of venous peak lactate, lactate-to-pyruvate ratio, and fold increase in lactate from baseline in aerobic exercise tests in patients with mitochondrial diseases, Intern Med Tokyo Jpn 61 (2022) 1939–1946, https://doi.org/10.2169/internalmedicine.8629-21.
[34] R.J. Dandurand, P.M. Matthews, D.L. Arnold, D.H. Eidelman, Mitochondrial disease. Pulmonary function, exercise performance, and blood lactate levels, Chest 108 (1995) 182–189, https://doi.org/10.1378/chest.108.1.182.
[35] T. Taivassalo, T.D. Jensen, N. Kennaway, S. DiMauro, J. Vissing, R.G. Haller, The spectrum of exercise tolerance in mitochondrial myopathies: a study of 40 patients, Brain J Neurol 126 (2003) 413–423, https://doi.org/10.1093/brain/awg028.
[36] T. Taivassalo, A. Abbott, P. Wyrick, R.G. Haller, Venous oxygen levels during aerobic forearm exercise: an index of impaired oxidative metabolism in mitochondrial myopathy, Ann. Neurol. 51 (2002) 38–44, https://doi.org/10.1002/ana.10027.
[37] T.D. Jensen, P. Kazemi-Esfarjani, E. Skomorowska, J. Vissing, A forearm exercise screening test for mitochondrial myopathy, Neurology 58 (2002) 1533–1538, https://doi.org/10.1212/wnl.58.10.1533.
[38] E. Hammarén, L. Rafsten, M. Kreuter, C. Lindberg, Modified exercise test in screening for mitochondrial myopathies–adjustment of workload in relation to muscle strength, Eur. Neurol. 51 (2004) 38–41, https://doi.org/10.1159/000074981.
[39] N. Løkken, S.V. Skriver, T. Khawajazada, J.H. Storgaard, J. Vissing, Plasma lactate responses during and after submaximal handgrip exercise are not diagnostically helpful in mitochondrial myopathy, Mitochondrion 60 (2021) 21–26, https://doi.org/10.1016/j.mito.2021.07.002.
[40] J. Bleistein, S. Zierz, Partial deficiency of complexes I and IV of the mitochondrial respiratory chain in skeletal muscle of two patients with mitochondrial myopathy, J. Neurol. 236 (1989) 218–222, https://doi.org/10.1007/BF00314503.
[41] A. Chan, R. Gold, S. Arp, K.W. Pflughaupt, K.V. Toyka, H. Reichmann, [Standardized bicycle ergometry test in mitochondrial myopathies. Indications, interferences and clinical parameters], Nervenarzt 69 (1998) 472–484, https://doi.org/10.1007/s001150050300.
[42] P.K. Pedersen, J.R. Nielsen, Absolute or relative work load in exercise testing–significance of individual differences in working capacity, Scand. J. Clin. Lab. Invest. 44 (1984) 635–642.
[43] Reproducibility of the lactate stress test - PubMed n.d. https://pubmed.ncbi.nlm.nih.gov/12822834/. (Accessed 26 June 2023).