European Journal of Physical and Rehabilitation Medicine 2024 Jun 05
DOI: 10.23736/S1973-9087.24.08187-5

SYSTEMATIC REVIEW

Checks and balances: a meta-analysis on the known-groups validity of functional postural control tests in children

Jorn OCKERMAN [1], Silke VELGHE [2], Anke VAN BLADEL [1], Edouard AUVINET [3], Jelle SALDIEN [4], Katrijn KLINGELS [2], Lynn BAR-ON [1], Evi VERBECQUE [2] *

[1]Unit of Pediatric Rehabilitation Research of Ghent, Department of Rehabilitation Sciences, Ghent University, Ghent, Belgium; [2]Rehabilitation Research Center (REVAL), Faculty of Rehabilitation Sciences, Hasselt University, Diepenbeek, Belgium; [3]Health Engineering Laboratory, Centre de Recherche et de Développement ECAM (CERDECAM), ECAM, Brussels, Belgium; [4]Media, Innovation and Communication Technologies, Department of Industrial Systems Engineering and Product Design, University of Ghent, Ghent, Belgium

*Corresponding author: Evi Verbecque, Rehabilitation Research Centre (REVAL), Faculty of Rehabilitation Sciences, Hasselt University, Agoralaan gebouw D, 3590, Diepenbeek, Belgium. E-mail: evi.verbecque@uhasselt.be

ABSTRACT

INTRODUCTION: Pediatric physical therapists commonly treat children with postural control deficits. Ideally, pediatric functional postural control tests should therefore be able to identify postural control deficits in children with various disorders. Despite a plethora of available tests, evidence for their validity – especially known-groups – remains scarce. This review aims to determine the known-group validity of available functional postural control tests to differentiate various pediatric pathological groups of different ages from their typically developing (TD) peers.
EVIDENCE ACQUISITION: PubMed, Web of Science and Scopus were systematically searched (last update: February 2023; PROSPERO: CRD42023408982). Forty case-control studies with a pathological pediatric sample (N.=1331) and TD peers (N.=1889) were included and selected for data-extraction and -analysis. Risk of bias was assessed using the SIGN checklist and level of evidence was scored using GRADE. Random-effect meta-analyses were performed to estimate pooled standardized mean differences (SMD) for the various test types and subclassified based on pathology and/or age.
EVIDENCE SYNTHESIS: When compared with TD peers, children with underlying pathologies performed significantly worse on pediatric functional postural control test batteries (SMD=-2.21), the Timed Up and Go Test and variants (SMD=2.30), the One Leg Stance test and variants (SMD=-2.14), while the Reach tests showed a smaller difference (SMD=-1.19). Subclassification within the meta-analyses showed that pathology was an influencing factor for the test batteries and the one leg stance test and variants. Age was an influencing factor for the reach tests. None of the included functional postural control tests exceeded a low level of evidence.
CONCLUSIONS: Pediatric functional postural control tests that assess multiple aspects of postural control (such as test batteries) seem to offer higher known-groups validity than single-task tests (*e.g.* reach tests). The underlying pathology has a larger impact on the validity of these tests than age. There remains an overall low level of evidence for the known-groups validity of pediatric functional postural control tests indicating the need for research with more homogenous groups and norm reference data.

(*Cite this article as:* Ockerman J, Velghe S, van Bladel A, Auvinet E, Saldien J, Klingels K, *et al.* Checks and balances: a meta-analysis on the known-groups validity of functional postural control tests in children. Eur J Phys Rehabil Med 2024 Jun 05. DOI: 10.23736/S1973-9087.24.08187-5)

KEY WORDS: Children; Postural balance; Reproducibility of results.

Introduction

Postural control, which helps maintain the body's upright position, move around freely, and protect an individual from falls and injury, is a fundamental prerequisite for independent mobility. Postural control deficits are prevalent in different pediatric disorders including, but not limited to: Developmental Coordination Disorder (DCD),[1]

Down Syndrome,[2] cerebral palsy (CP),[3] vestibular conditions,[4] etc. However, the prevalence of the postural control deficits differs depending on the type of pathology with ranges between 48% for children with CP[3] and up to 87% for children with DCD.[1] Depending on the underlying pathology, a variability in the severity and developmental progress of postural control deficits has been reported.[5] These deficits can hinder daily tasks, playground activities, school performance, and learning new motor skills, potentially affecting the child's self-esteem and social interaction at a socio-emotional level. To estimate the severity of the postural control deficits and their impact on overall functioning, and evaluate the efficacy of therapeutic interventions, adequate assessment tools are required.

Therapists should use standardized assessment tools to objectively evaluate a child's health and establish a physical therapy diagnosis by identifying movement system impairments.[6] By allowing therapists to objectively assess clinically relevant treatment-induced changes, they also facilitate the development of a personalized intervention based on evidence, while also enabling ongoing adjustments to optimize health outcomes for the patient.[6-8]

As defined by Horak,[9] postural control assessment instruments should be able to identify the presence of a postural control deficit and determine the underlying cause. According to Mancini & Horak,[10] postural control can be assessed using a functional approach, whereby predefined motor tasks are rated on ordinal scales often incorporating performance times or number of correct repetitions. Although various functional postural control tests are available[11, 12] they, and standardized measurement instruments in general, are still not systematically implemented in clinical practice. While a lack of time and financial or organizational incentives are reported as primary barriers to systematic use, recent studies also highlighted a lack of knowledge and skill as potential hurdles clinicians face.[13-15] When it comes to postural control tests, no conclusive clinical guidelines exist, partially due to the limited evidence regarding the psychometric properties of these tests. To select the most suitable functional postural control test for their patients, they need to know to what extent a test is generalizable and applicable in various populations, *i.e.* external validity. Moreover, such a test should be able to quantify clinical problems. If a child or their parents request help for a specific postural control problem (*e.g.* frequent falls) and a postural control test is unable to identify any deficits or quantify them appropriately, that measure would simply be unsuitable for use in that specific case.

A recent systematic review on the psychometric properties and clinical utility of functional postural control tests showed that comprehensive tests which cover a broader construct of postural control (*e.g.* the Pediatric Balance Scale [PBS], the Fullerton Advanced Balance Scale, etc.) are preferred over tests that only account for one aspect of postural control (*e.g.* the One-Leg Stance, the Clinical Test of Sensory Interaction on Balance, etc.).[11] While researchers found sufficient levels of reliability, evidence was rated very low or low for most tests, except for the Early Clinical Assessment of Balance (ECAB), Four Square Stepping Test (FSST) and Balance Error Scoring Scale (BESS) where medium levels of evidence were found. Measures of responsiveness and validity for these functional postural control tests were found to be insufficiently investigated. Although adequate responsiveness is important to measure treatment effects and assess the defined treatment plan, a test's validity is even more important, since it provides insights into whether the test actually measures what it intends to measure. One of the vital types of construct validity is known-groups validity, which assesses the degree to which an instrument can demonstrate different scores for groups known to vary on the variable being measured.[16] In a clinical context this translates to the degree to which a test is able to identify a child with postural control deficits based on their difference to age- or norm-referenced data. Known-groups validity can therefore be used to distinguish between age groups, pathological from typically developing (TD), or a combination of both. The descriptive, systematic review by Johnson *et al.*, reported a poor to very poor level of evidence for the known-groups validity of most functional postural control tests.[11] However, the achieved known-groups validity per available test was not quantified via a meta-analysis of pooled results.

Moreover, postural control is known to develop with increasing age, starting from infancy, and continuing through early childhood up to adolescence.[17] The exact time at which a child reaches mature levels of postural control is still under discussion and depends on the type of task and the outcome variable used. For instance, some posturography studies suggest maturation to occur within the age range between 7 and 10 years old,[18, 19] while other studies present evidence of an ongoing maturation process during puberty (13-15 years old), possibly reaching to the end of adolescence.[17, 20, 21] Functional assessment tools tend to have ceiling effects earlier, *e.g.* maximal performances on the PBS have been reported by the age of seven,[22] standing on a foam pad with eyes closed for 30 seconds by the age of eight,[23] or reaching forward as far as possible by the age

of 12.[24] A recent study exploring age-related differences in performance on functional postural control tests, indicated that particularly above the age of 13, all the test items of the mini Balance Evaluation Systems were fully mastered except for the TUG with a dual task and the compensatory backward stepping response.[25] Due to this non-linear development of postural control and the task-specificity of the developmental progressions, the ability to perform functional tasks that require a large amount of postural control, will also vary with age.

Abnormal developmental patterns could further influence this variability in the acquisition of motor skills. A summary categorization of abnormal development lists the following four categories: 1) normal development, defined by a median and normal range; 2) regressive development, whereby gained skills are lost over time; 3) plateau effect, where no progression is possible past a certain point; and 4) slow but steady development whereby the gap between normal and abnormal development becomes larger with increasing age and thus more apparent over time.[26] Most functional postural control tasks are relatively simple and more prone to become obsolete once a certain developmental stage has been reached. While some patterns of abnormal development (plateau and regressive) are easily picked up by these simple tests, children experiencing a slower development could still remain undiagnosed when a simple test lacks the required level of known-groups validity.

Age-specific normative data are available for several, but not all, pediatric functional postural control tests.[11] These age-specific normative data indicate that with increasing age, performance in TD children changes and that these developmental changes should be considered during assessment that aims to identify postural control deficits. The use of such age-specific normative data is therefore a cornerstone in the identification of postural control deficits in children with an underlying pathology. Despite the availability of these norm data, they have, to the best of our knowledge, not yet been used to identify postural control deficits in children with underlying pathologies. Rather, authors explore the extent to which the functional performance of a clinical group (*e.g.* children with CP) differs from age- (and sex-) matched TD peers (*i.e.* case-control designs). Such group-level results are based on the data of subjects of different ages, making a direct comparison to age-specific normative data impossible. Nevertheless, the insights gained through these case-control study designs can add to our knowledge of known-groups validity of the functional postural control tests.[21] Therefore, a first step to determine the generalizability and clinical applicability of the functional postural control tests for which normative data already exist, is to pool the results from multiple case-control studies and establish the known-groups validity of these tests in larger samples.

This review aims to investigate the known-groups validity of pediatric functional postural control tests by performing a systematic review and meta-analysis of study results where pediatric pathological groups were compared with TD peers. Furthermore, this review also aims to determine how pathology and/or age affect the known-groups validity of functional postural control tests.

Evidence acquisition

Protocol registration

The present review was performed following the Preferred reporting Items for Systematic reviews and Meta-analyses (PRISMA) guidelines[27] and registered in PROSPERO (CRD42023408982).

Study design

A systematic review and meta-analysis of case-control studies implementing specific functional postural control tests in various pediatric populations was performed.

Search strategy

The electronic database Medline (PubMed interface), Scopus and Web of Science were searched using the PICOs-method (pediatric patient populations (P), functional assessment tools for postural control (I), TD children (C), numeric values (means and SDs/medians and interquartile range) characterizing performances (O), case-control design (S)). We wanted to explore known-groups validity for all the test of which normative data are available. As such, we first searched the literature for new publications on normative data of pediatric functional postural control tests with the search string reported by Johnson *et al.*, supplemented with synonyms covering "reference values" and "normative data." Since no new assessment instruments were identified by this search, the final search string included all identified functional postural control tests[11] separated by the Boolean term "OR," followed by a combination of free-text and MeSH-terms representing the concept "children" (Supplementary Digital Material 1: Supplementary Text File 1). For a detailed description of the selected functional postural control tests see the previous review.[7] The last update of the search was conducted on the 6th of February 2023.

Eligibility criteria

Studies were included in this review if they: 1) reported for their study sample a mean age below 13. This cut-off was determined based on the fact that pediatric functional postural control test often involve simple motor tasks for which TD children older than 13 are able to obtain maximum scores.[22-25] 2) Included both a pathological population and a TD control group; 3) described a performance-based assessment tool that evaluated postural control; 4) reported numerical distributions (mean and standard deviation [SD] or median and interquartile range [IQR]) of the postural control tests; and 5) were written in English, Dutch, French, German, or Spanish. Articles were excluded if: 1) their samples consisted of pediatric elite-athletes or children with acute sport-related traumatic brain injuries (*e.g.*, concussion); 2) reported any type of intervention; 3) insufficiently reported on the numerical distributions (*e.g.* only mean, median, graph); 4) or reported digital outcomes involving specific hardware (*e.g.* posturography, pressure mats, accelerometry etc.). This latter criterion was added as, although such hardware may be useful in quantifying postural control deficits, they are beyond the scope of this review that aimed to summarize the evidence of commonly applied functional postural control tests.

Study selection

The selection process was performed by three of the authors (JO, SV, EV), each screening two thirds of the articles to ensure every article was screened by two independent reviewers. The reviewers initially screened titles and abstracts using the following order of exclusion criteria: population, study design, outcome measures, and language. Screening of the remaining full texts was performed by the same reviewers. Discrepancies at any stage were discussed with the second reviewer and when consensus could not be reached, with the third.

Risk of bias assessment

Methodological quality of the individual studies was evaluated using the Scottish Intercollegiate Guidelines Network (SIGN) Methodology Checklist for case-control studies.[28] After all reviewers completed a trial sample of four studies, all remaining studies were independently scored by two reviewers. In accordance with the guidelines of the SIGN checklist, reviewers evaluated aspects of each study with the options "yes," "no" or "can't say." After a consensus meeting with both reviewers, all articles received an overall score referring to Low, Acceptable or High Quality. Lower than six positive responses resulted in a Low-quality score, while studies scoring more than eight positive responses received a High-quality score. Articles with more than two negative scores on statements evaluating the comparability of populations, definition of cases, quality of measures, or analysis of confounders were excluded from this review.[28]

Data extraction and synthesis

Information from each study was extracted into an excel sheet, containing information about: the general pathological group (*i.e.* genetic/neurological/neurodevelopmental/metabolic/oncological/ sensory disorder); pathology-specific diagnosis, severity, and classification levels; the performance-based test (name, protocol specifics, overarching test type category); and the biometric data (mean age, weight, height, and BMI) and outcome scores (mean or median score and standard deviation or 95% confidence interval [CI]) of participants in both the case and control groups.

Data-analysis

Statistical Package for Social Sciences (SPSS) version 28.0 software was used to perform random-effects meta-analyses to estimate pooled standardized mean differences (SMD) for the outcomes. At least three relevant articles reporting on one specific functional postural control test were required to enable the meta-analysis. Random-effects meta-analyses were chosen to incorporate the expected random variation in the effect across the studies into the pooled estimates. All results are presented with 95% CI.

Clinical (variability in the participants, assessments, and outcomes of the studies), methodological (study design, risk of bias) as well as statistical diversity were considered when assessing heterogeneity. The statistical heterogeneity was calculated using the I^2 test to describe the percentage of variation across studies which is due to heterogeneity. When heterogeneity values were ≥50%, subgroup analysis was considered accounting for the clinical diversity in the following sequence: underlying pathology or age of the participants.

Level of evidence

The quality of evidence of each pediatric functional postural control test included in this study was assessed using the Grading of Recommendation, Assessment, Development, and Evaluation (GRADE) method.[29] Each test underwent a comprehensive evaluation of evidence certainty,

considering factors such as risk of bias, precision, consistency, and directness. Precision was evaluated based on the 95% CI magnitude, where smaller intervals signified higher precision, in combination with an adequate sample size. Consistency was determined by analyzing variance in point estimates across studies, overlap of 95% CIs, and heterogeneity (I^2 value). Directness was assessed by examining differences in population characteristics and variations in outcome measures.

Each evaluated aspect (risk of bias, precision, consistency, and directness) contributed to an overall judgment, categorized as follows: 1) high quality, indicating that further research is unlikely to substantially alter our confidence in the effect estimate; 2) moderate quality, suggesting that future research is likely to significantly influence our confidence in the estimate and may lead to changes; 3) low quality, signifying that future research is highly likely to impact our confidence in the estimate and is likely to result in changes; 4) very low quality, reflecting a high degree of uncertainty about the estimate. Two reviewers (JO and EV) independently assessed the quality of the evidence assessment, and any discrepancies were resolved through discussion.

Results

Information sources and search strategy

The literature search resulted in a total of 1612 unique articles. After screening on title and abstract, the full texts of the remaining 95 studies were screened for eligibility. Eventually, 42 studies were selected to include in this systematic review (Figure 1). Two studies were excluded based on the SIGN checklist, leaving 40 studies included for data-extraction and -analysis.

Risk of bias assessment

Based on the risk of bias assessment, 19 studies showed acceptable quality,[30-47] 22 studies were rated as low quality[2, 48-66] and two studies[67, 68] were excluded. Details on the risk of bias assessment are presented in Table I.[2, 30-70] All studies addressed an appropriate and clearly focused research question, measured the exposure status in a standardized, valid, and reliable way and provided confidence intervals. The majority of the studies clearly defined cases and differentiated them from controls (67.5%) and took the main confounding factors into account in the design

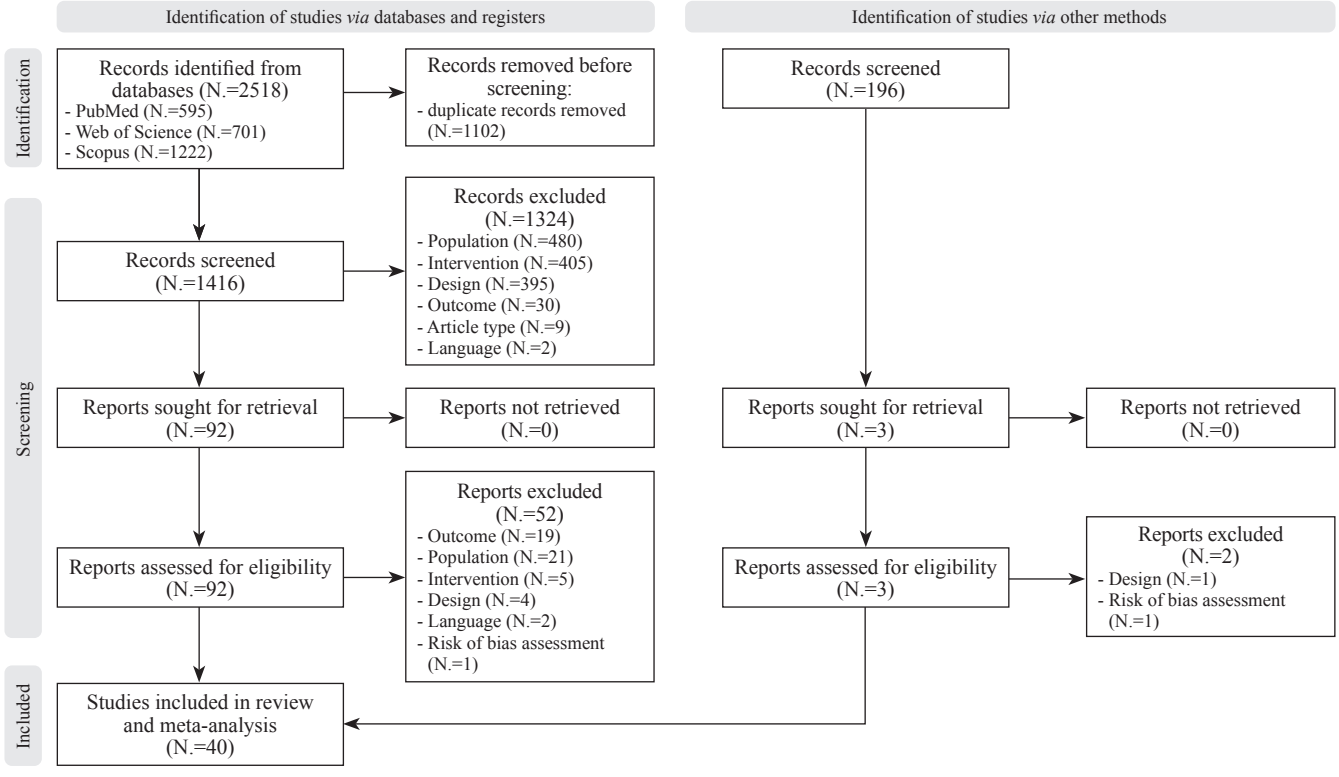


Figure 1.—Flow diagram of the selection process.

TABLE I.—*Risk of bias assessment.*[2, 30-70]

Authors/SIGN checklist	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	1.10	1.11	2.1
Abdul Rahman[30]	+	?	+	NR	-	+	-	?	+	+	+	Acceptable
Alkan[48]	+	?	+	NR	-	+	-	?	+	-	+	Low
Bas[49]	?	-	?	NR	-	+	+	?	+	-	+	Low
Bricout[31]	+	+	+	NR	-	+	+	?	+	+	+	Acceptable
Buker[32]	+	+	+	NR	-	+	+	?	+	+	+	Acceptable
Coffey[50]	+	-	?	NR	-	?	?	?	+	+	+	Low
De Kegel[52]	+	?	?	NR	-	+	-	?	+	+	+	Low
De Kegel[51]	+	?	?	NR	-	+	-	?	+	-	+	Low
Deforche[53]	+	+	?	NR	-	?	?	?	+	+	+	Low
Dewar[33]	+	?	+	NR	-	?	+	?	+	+	+	Acceptable
Dik[34]	+	-	+	NR	-	+	+	?	+	+	+	Acceptable
Franjoine[54]	+	-	?	NR	-	?	-	?	+	+	+	Low
Goldman[55]	+	+	?	NR	-	?	-	?	+	-	+	Low
Graff[68]	+	?	-	NR	-	-	?	?	?	-	+	Excluded
Haibach[69]	+	-	?	NR	-	-	-	?	+	?	+	Low
Huang[70]	+	-	+	NR	-	?	-	?	+	+	+	Low
Jain[2]	+	?	?	NR	-	+	-	?	+	+	+	Low
Jayakaran[35]	+	?	+	NR	-	+	+	?	+	+	+	Acceptable
Kalyani[36]	+	?	+	NR	-	+	+	?	+	+	+	Acceptable
Karakoc[37]	+	?	+	NR	-	?	+	?	+	+	+	Acceptable
Katz Leurer[58]	+	?	?	NR	-	+	-	?	+	+	+	Low
Katz Leurer[56]	+	?	?	NR	-	+	-	?	+	+	+	Low
Katz Leurer[57]	+	?	?	NR	-	+	-	?	+	+	+	Low
Kaya[38]	+	-	+	NR	-	+	+	?	+	-	+	Acceptable
Kembhavi[59]	+	-	+	NR	-	+	-	?	+	-	+	Low
Leizerowitz[39]	+	-	+	NR	-	+	-	?	+	+	+	Acceptable
Lukacs[40]	+	?	+	+	+	+	?	?	+	+	+	Acceptable
Marchese[60]	+	?	?	NR	-	+	?	?	+	+	+	Low
Melo[41]	+	+	-	NR	-	+	?	?	+	+	+	Acceptable
Melo[42]	+	+	+	NR	-	+	+	?	+	+	+	Acceptable
Mohammed[61]	+	?	?	NR	-	+	?	?	+	+	+	Low
Newman[43]	+	?	?	+	-	+	-	?	+	+	+	Acceptable
Newman[62]	+	?	?	NR	-	+	-	?	+	+	+	Low
Nicolini-Panison[44]	+	-	?	+	-	-	+	?	+	+	+	Acceptable
Pace & Bricout[45]	+	?	+	NR	-	+	?	?	+	+	+	Acceptable
Promsorn & Taweetanalarp[63]	+	-	?	NR	-	?	-	?	+	+	+	Low
Roostaei[66]	+	?	-	NR	-	?	-	?	?	-	+	Low
Sanz-Santiago[46]	+	+	-	+	+	+	?	?	+	+	+	Acceptable
Soares[47]	+	-	-	+	-	+	?	?	+	+	+	Acceptable
Tsiros[64]	+	?	+	NR	-	?	?	?	+	+	+	Low
Walowska & Bolach[67]	+	-	?	NR	-	-	-	?	+	-	+	Excluded
Zaino[65]	+	-	?	NR	-	?	-	?	+	+	+	Low

1.1 The study addresses an appropriate and clearly focused question; 1.2 The cases and controls are taken from comparable populations; 1.3 The same exclusion criteria are used for both cases and controls; 1.4 What percentage (%) of each group (cases and controls) participated in the study?; 1.5 Comparison is made between participants and non-participants to establish their similarities or differences; 1.6 Cases are clearly defined and differentiated from controls; 1.7 It is clearly established that controls are non-cases; 1.8 Measures will have been taken to prevent knowledge of primary exposure influencing case ascertainment; 1.9 Exposure status is measured in a standard, valid and reliable way; 1.10 The main potential confounders are identified and taken into account in the design and analysis; 1.11 Confidence intervals are provided; 2.1 How well was the study done to minimize the risk of bias or confounding? "+" yes, the study does this; "?" can't say whether the study does this ; "-" no, the study does not do this; NR: not reported.

and/or analysis (77.5%). Most studies were unclear about whether cases and controls were taken from comparable populations (80%), the percentage of cases and controls that participated in the study relative to the number of invited participants (85%) and whether measures had been taken to prevent knowledge of primary exposure (blinding of assessors) (100%). The majority of the studies did not make a comparison between participants and non-participants (92.5%), nor did they clearly establish whether controls were non-cases (70%).

Participants

Postural control was assessed in a total of 1331 children with underlying pathologies (mean age [SD]: 8.68 y [3.72]) and 1889 controls (mean age [SD]: 9.10 y [4.04]). We distinguished seven overarching groups related to pathology: genetic syndromes (CHARGE Syndrome,[69] Down Syndrome,[2, 36, 38, 44, 63] genetic variant 16p11.2[55]), metabolic disorders (diabetes type I,[40] overweight[53] and obesity[64]), neurodevelopmental disorders (attention deficit hyperactivity disorder,[32, 61] autism spectrum disorder[31, 45, 50]), neurological disorders (CP,[33, 39, 47, 54, 56, 59, 65, 66, 70] acquired/traumatic brain injury,[30, 39, 43, 56-58, 62] Duchenne's muscular dystrophy [DMD][48]), oncologic disorders,[60] respiratory disorders (cystic fibrosis,[34] asthma[46]) and sensory disorders (visual impairment,[49] hearing impairment[37, 41, 42, 51, 52]). A detailed description of the cases and controls participating in the original studies can be consulted in Supplementary Digital Material 2, Supplementary Table I.

Outcome measures

The outcome measures were grouped into seven overarching categories related to the test type: test batteries (Pediatric Balance Scale [PBS],[2, 30, 35, 36, 38, 42, 54, 61, 69] the Berg Balance Scale [BBS],[59] Balance Evaluation Systems test for children [Kids-BESTest][33]), reach tests (sitting RT,[48] forward RT,[2, 37, 47-49, 56-58, 65, 70] lateral RT,[2, 58, 70] pediatric RT[34] and multidirectional RT[30, 63]), Timed Up and Go Test (TUG) and variants (classic TUG,[37-39, 46, 55-58, 60, 64-66] extended versions[48] and TUG with dual task[43, 62, 66]), one leg stance and variants (Timed OLS,[31, 37, 51, 52, 65] Flamingo Test,[31, 32, 40, 45] Stork Balance Test[50] and Y-Balance Test[32]), sensory perturbation tests (Tandem stance,[31] BESS,[41] Standing on a beam/eyes open/eyes closed[31, 53]), tandem walking (backward[31] and forward[31, 53]) and other timed measures (Four-square stepping test[39] and Timed Up and Down Stairs[46, 65]). The detailed description of these tests can be found in Appendix D of the systematic review by Johnson *et al.*[11]

Differences in performances between cases and controls

Differences between cases and controls were analyzed per functional postural control test and are presented as such in the next paragraphs. Figure 2 provides a summary of the pooled SMDs and subgroup analyses where applicable.

Test batteries

In total, 13 studies investigated performances on test batteries in children with underlying pathologies (N.=518) and compared them to TD peers (N.=617). Eleven studies applied the PBS[2, 30, 35-38, 42, 49, 54, 61, 69], one study the BBS,[59] and one the Kids-BESTest.[33]

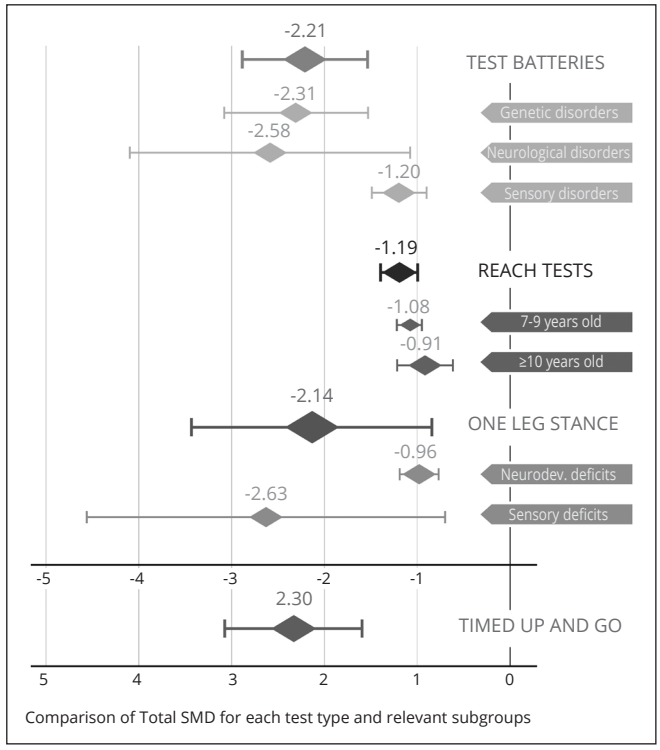


Figure 2.—Summary of the pooled standardized mean difference between cases and controls on functional postural control tests.

Children experiencing motor difficulties performed 2SDs poorer on postural control test batteries than their TD peers (SMD=-2.21, 95% CI=[-2.90;-1.52], P<0.001, I^2=96%; Supplementary Digital Material 3: Supplementary Figure 1). The high heterogeneity suggests the need for subclassification, taking into consideration the group of underlying deficits.

Overall, children with genetic syndromes perform significantly poorer than their TD peers on the PBS (SMD=-2.31, 95% CI=[-3.09;-1.52], P<0.001, I^2=79%).[2, 36, 69] The difference seems to be more marked in younger children (below age 9), but not enough studies were available to explore this with a subgroup analysis (Supplementary Digital Material 4: Supplementary Figure 2). Children with neurological disorders (CP[33, 54, 59] and traumatic brain injury[30]) also perform significantly weaker on the PBS, BBS and Kids-BESTest than their TD peers (SMD=-2.58, 95% CI=[-4.10;-1.06], P<0.001, I^2=98%). Yet, the level of motor functioning among children with neurological disorders seems to play an important role in the distribution of the test results, regardless of the applied test battery (Sup-

plementary Digital Material 5: Supplementary Figure 3).

Children experiencing sensory deficits[35, 37, 42] perform significantly poorer than their TD peers (SMD= -1.20, 95% CI=[-1.50;-0.89], P<0.001, I^2=13%). See Supplementary Digital Material 6: Supplementary Figure 4 for details on the subgroup meta-analysis.

Reach tests

Twelve studies investigated differences in postural control between children with various underlying pathologies and their TD peers using functional reach tests.[2, 34, 37, 47-49, 56-58, 63, 65, 70] A large variety of reach tests was identified: the sitting reach test,[48] the classic functional reach test in the forward[2, 37, 47-49, 56-58, 65, 70] and lateral direction,[2, 58, 70] the multidirectional reach test[63] and the pediatric reach test.[34] The reach performance in children with underlying pathologies was on average 1.2 SD lower than those of their TD peers (SMD=-1.19, 95% CI=[-1.41;-0.98], P<0.001, I^2=65%) . In contrast to the test batteries where the pathological groups tended to impact the magnitude of the difference relative to TD peers, this was not the case for the reach tests (Supplementary Digital Material 7: Supplementary Figure 5). Age on the other hand seems to play a crucial role. When children are grouped according to their mean group age into 7-9 and ≥10 years old, heterogeneity among studies reduces. In the age band 7-9 years old (Supplementary Digital Material 8: Supplementary Figure 6), children with underlying motor deficits, reach approximately 1.1 SD less far when compared to their TD peers (SMD=-1.08, 95%CI=[-1.23;-0.94], P<0.001, I^2=0%). In the children aged 10 and older, similar results are noted (SMD=-0.91, 95%CI=[-1.22,-0.6], P<0.001, I^2=22%) (Supplementary Digital Material 9: Supplementary Figure 7).

The TUG and variants

The TUG was applied in 17 studies.[37-39, 43, 44, 46, 48, 49, 55-58, 60, 62, 64-66] Children with underlying pathologies need 2.3 SD more time to execute the TUG compared to their TD peers (SMD=2.30, 95%CI=[1.51; 3.10], P<0.001, I^2=97%). The high amount of heterogeneity among the included studies, suggests the need for subclassification. However, none of the available grouping variables (pathological group, age group and methodological quality) resulted in less heterogeneity (Supplementary Digital Material 10: Supplementary Figure 8).

One leg stance and variants

In 11 studies the difference in one leg stance performance was investigated between children with underlying pathologies and TD peers.[31, 32, 37, 40, 45, 49-53, 65] Children with various pathologies on average perform approximately 2 SDs poorer than their TD peers (SMD=-2.14; 95%CI=[-3.45;-0.83], P<0.001; I^2=99%), (Supplementary Digital Material 11: Supplementary Figure 9). The large heterogeneity seems to result from the different pathological conditions. Children with neurodevelopmental disorders perform significantly weaker than their TD peers (SMD=-0.96; 95%CI=[-1.18; -0.74], P<0.001; I^2=0%; Supplementary Digital Material 12: Supplementary Figure 10). Among the group of children with sensory deficits, clear differences arise from the type of sensory deficit (SMD=-2.63; 95%CI = [-4.57;-0.69], P<0.001; I^2=99%), but further subclassification was not possible due to too few available data (Supplementary Digital Material 13: Supplementary Figure 11).

Sensory perturbations

Three studies reported on various postural control tests tapping into sensory orientation and/or reweighting: the modified Clinical Test of Sensory Interaction on Balance,[49] the Balance Error Scoring System[41] and different sustained postures.[31] The mean SMD at study level varied between -0.47[31] to -3.10[49] in favor of the control group.

Tandem walk

Two studies explored differences in postural control in overweight children[53] and children with ASD[31] compared to their TD peers using the tandem walk backward[31] and forward.[31, 53] Yet, TD children tend to make significantly fewer errors than children with ASD (P<0.01)[31] and take more steps than overweight children (P<0.001).[53]

Other timed measures

Three studies applied timed measures other than the TUG or OLS variants.[39, 46, 65] Two studies used the Timed Up and Down Stairs test (TUDS), one of which investigated children with CP[65] and one children with asthma[46] and one study administered the Four Square Stepping Test (FSST).[39] At an individual study level, the clinical groups need significantly more time to complete the TUDS (P<0.02)[65] and the FSST (P<0.001)[39] compared to their TD peers.

Level of evidence

There was low evidence for the use of test batteries and very low evidence for the RT, TUG and variants, and OLS and variants, due to inconsistency of results, impression, and risk of bias. Details are listed in Table II.

TABLE II.—*Level of evidence.*

Functional postural control test groups	Study Limitations	Inconsistency of result	Indirectness of evidence	Imprecision	Quality of the evidence (GRADE)
Test batteries (PBS and Kids-BESTest)	↓	=	=	↓	⊕⊕⊖⊖
Reach tests	↓↓	↓	=	=	⊕⊖⊖⊖
TUG and variants	↓↓	↓	=	↓	⊖⊖⊖⊖
OLS and variants	↓	↓↓	↓	↓↓	⊖⊖⊖⊖

↓GRADE score downgraded by one point; = No impact on GRADE score.
⊕ Point on final GRADE score awarded; ⊖ Point on final GRADE score not awarded and suggested representations of the quality of evidence; Symbol ⊕⊕⊕⊕ = high; ⊕⊕⊖⊖ = moderate; ⊕⊖⊖⊖ = low, and ⊕⊖⊖⊖ or ⊖⊖⊖⊖ = very low
PBS: Pediatric Balance Scale; TUG: Timed Up and Go Test; OLS: One Leg Stance.

Discussion

The aim of this systematic review was to gain a better understanding of the clinical value of functional postural control tests in pediatric populations by determining the known-groups validity of these tests when comparing children with pathologies and TD peers taking type of pathology and age into consideration. We identified 40 studies that met the criteria, thereby covering 22 different tests. We clustered and analyzed these tests in seven overarching categories. Of these categories, only the test batteries, the reach tests, the TUG and variants, and the OLS and variants yielded a sufficient number of studies to perform meta-analyses. Our results show that while some pediatric functional postural control tests are able to differentiate children with underlying pathologies from TD children, some important considerations should be kept in mind.

Known-groups validity

In general, the test batteries, TUG and OLS show the most promising known-groups validity as they show large pooled SMDs between cases and controls (Figure 2). Norm-referenced scales, such as the Movement Assessment Battery for Children, Second Edition[71] and the Bruininks-Oseretsky Test of Motor Proficiency, Second Edition,[72] use percentile rank scores to interpret performances. In such tests, scores at or below the 15-17th percentile (*i.e.* 1 SD below the mean) indicate a risk for poor performance, and a score at or below the 2-5th percentile rank indicates a definite problem.[71, 72] When using a similar interpretative method for the data collected in this review it can be concluded that specifically for the test batteries, the TUG and variants, and the OLS and variants (Figure 3): 1) more than 50% of all the cases perform below the 5th percentile rank of the control group, indicating a definite postural control problem and 2) more than 85% of all the cases perform below the 15th percentile. This implies that only a small portion (<15%) of the case group performs within the broad 'normal' range (up to 1 SD below the mean) which is in line with the expected clinical variation in postural control deficits in children with different underlying pathologies.[1-4] Contrarily, the reach tests show a much smaller difference between cases and controls (Figure 3), with approximately 50% of the cases performing similarly to TD children. Although the characteristics of the children performing the different functional postural control tests are similar (genetic, neurological, respiratory, and sensory disorders), the functional reach tests do not seem to identify the children with underlying postural control deficits to the same extent as the test batteries. This implies that this isolated task is probably insufficiently sensitive and that the assessment of multiple aspects of postural control (as is done in test batteries) is required to distinguish between groups.

Although the mean SMD promotes the use of the test batteries (the TUG and variants, and the OLS and variants), these results need to be interpreted cautiously. The very large 95% CIs of the SMDs indicate large imprecision, which is evidenced by the high amount of heterogeneity among studies (I^2>50%). Subclassification based on the specific overarching pathology groups (*e.g.* genetic or neurological disorders) showed that the children with sensory disorders performed significantly lower than their age-matched controls on the test batteries, but the magnitude was only±1 SD below the mean of the control group. Therefore, while subclassification reduced the imprecision and the heterogeneity, it also reduced the magnitude of the difference between specific cases and controls. A similar result was found for children with neurodevelopmental disorders regarding the OLS and variants. These findings can be explained by the fact that not all children with these pathologies present with a postural control deficit.[1-4] Defining both homogenic groups and mapping the extent to which these children actually experience postural control deficits, can have a major impact on the known-groups validity of the test which in turn, affects its sensitivity and generalizability.

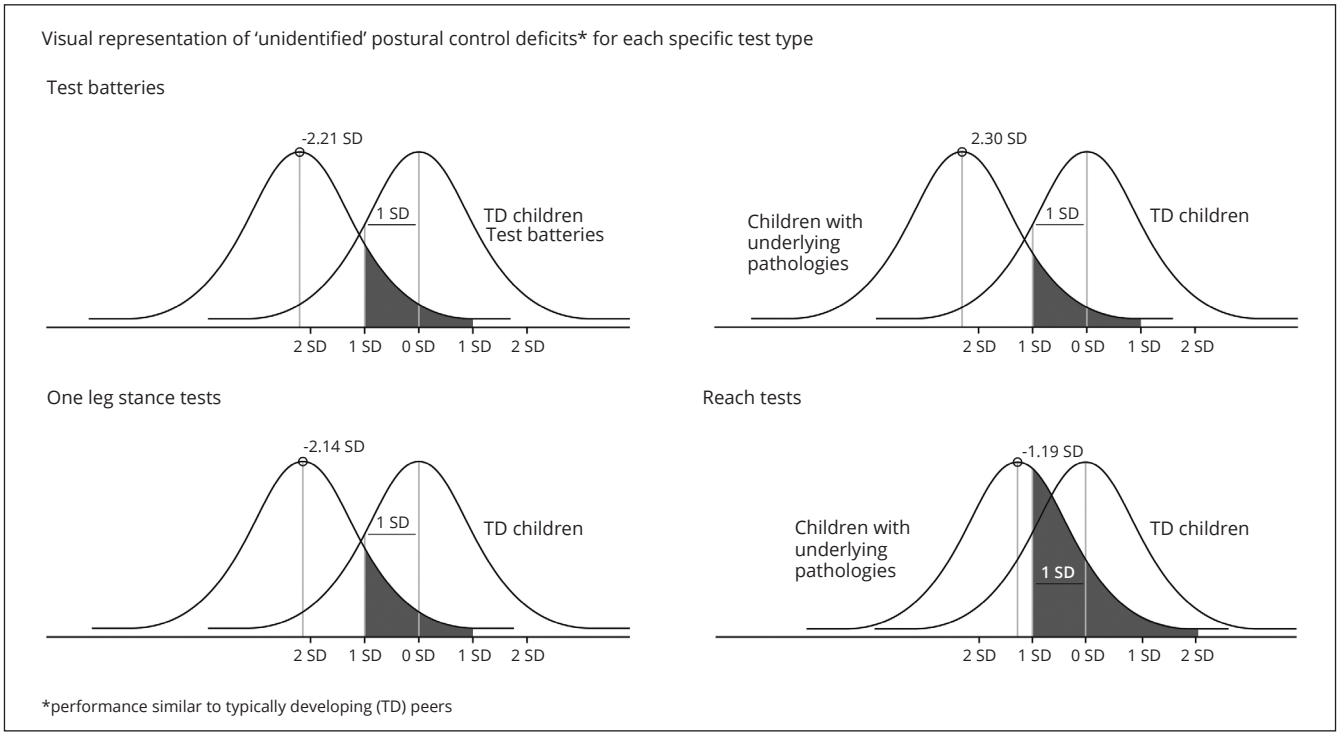


Figure 3.—Visual representations of 'unidentified' postural control deficits for the test batteries, TUG tests, OLS tests and Reach Tests.

Influencing factors

As hypothesized, both pathology and age, and the interaction between them, affected the observed differences in known-groups validity, but the degree to which they did, was specific to each individual category of functional postural control test.

Due to the non-linear development of postural control, it stands to reason that age-related factors would also affect the difference in performance between children with underlying pathologies and their TD peers. However, the results of our meta-analyses seem to indicate that these expected age-related effects are not a singular explanation for the reported heterogeneity of results. Only for the RT category, a stark decrease in heterogeneity was seen once a subclassification based on age (mean group age 7-9 years and ≥10 years) was applied. While both the younger and older groups showed similar effect sizes (SMD=-1.08 and SMD=-0.91 respectively) the level of heterogeneity for each group dropped significantly (I^2=0% and I^2=22% respectively). Reach distance is known to be related to age, anthropometric factors such as height, arm length, and weight, and to base of support.[24, 73-77] Since growth and development continues through childhood regardless of any underlying pathology, age may be a better explanatory variable than pathology for these specific tests. Interestingly, the test batteries also include an item with a reach task, but their scoring is done based on an ordinal rating scale instead of the reach distance, and the total is a combination of multiple items, which, all together, may conceal the relationship with age. Furthermore, although the PBS consists of tasks that, in TD children are fully controlled by the age of seven,[22] age is not a primary factor to explain the imprecision of these tests when comparing children with varying underlying pathologies to their TD peers (Figure 2). As such, it seems that multiple tasks covering several aspects of postural control and the underlying pathology has a larger impact on the magnitude of the difference than age.

Similarly to the RTs, the OLS and variants also emphasize a single aspect of postural control, *i.e.* anticipatory postural adjustments. Among TD children, the time a child can maintain the OLS position increases with age.[78] The included studies mainly reported the OLS in children with more mild pathologies such as neurodevelopmental and sensory disorders. This stresses the importance of considering task-complexity when discussing appropriate pedi-

atric functional postural control tests. Indeed, the severity level of a child's motor disability (*e.g.* Gross Motor Function Classification System or 'GMFCS' levels in CP) impacts the magnitude of the differences between cases and controls. When we stratify the results based on the general motor classification, we note that cohorts of children with higher GMFCS levels perform significantly worse on test batteries than their peers with a lower classification (Supplementary Digital Material 3). Unfortunately, due to the paucity of mean data for the various GMFCS levels, it was impossible to perform a meta-analysis based on this subclassification.

Our meta-analysis also suggested an interaction between the effects of age and overarching pathological group. For example, the heterogeneity of the PBS analysis was reduced when stratifying the results for both age and pathology. This interaction of both factors should be kept in mind when interpreting the results of functional postural control tests. While the development of motor skills in TD children is already non-linear and varied, this maturation process in children with an underlying pathology shows even more variation.[21] Furthermore, we can assume that the development of motor skills also differs from one pathological group to another. A child with CP will be faced with different difficulties at different points in their development than a child with DCD, hearing loss or a genetic disorder. When we consider the principles of the Dynamic Systems Theory, these variations in movement development will result in different patterns and strategies to perform specific motor tasks.[21] While task complexity clearly impacts performance and improves with age in TD children, this may not be the case in a child whose motor development adheres to an abnormal pattern (*i.e.* slow but steady, plateau or regression). Hence, the task choice during assessment will depend on whether they are relevant for that particular child's motor repertoire. For instance, the PBS can provide valuable information for a 10-year-old child with a central neurological deficit that is unable to walk. Hence, despite the fact that the PBS has a ceiling effect for TD children starting from the age of seven, children who have a plateau in their motor development that occurs before this motor age, still benefit from this type of assessment. Visa versa, a 10-year-old child with a mild motor deficit (slow but steady developmental pattern) who experiences postural control deficits during physical activity classes, is likely to have no difficulties with the PBS. Since these children adhere more closely to typical developmental patterns, performance variability on simple tests might not have sufficient known-groups validity to allow proper identification. More complex tasks (*e.g.* the Y-balance or the Stork balance test) might therefore be more suitable.

With all this in mind, it might be necessary for clinicians to have access to usable and relevant reference data, specific to either the age or developmental stage of the patient and their underlying pathology (if present). To reach acceptable levels of external validity, functional postural control tests and especially their associated reference data might need to be re-evaluated for them to be able to account for these influencing effects. Unfortunately, the current data do not yet support specific recommendations for age or pathology-specific normative data of pediatric functional postural control tests.

In summary, we can conclude that test batteries have the advantage of assessing the construct of postural control more extensively while boasting large SMDs between pathological groups and TD children, despite low levels of evidence. However, the generalizability of the test batteries seems limited to the regressive (*e.g.* DMD) and plateaued developmental patterns (*e.g.* CP). The strength of the TUG and variants is the inclusion of multiple activities of daily living in the assessment (standing up, sitting down, walking, and turning). However, these tests are not as comprehensive as the test batteries and show a very low level of evidence which limits their generalizability, despite larger SMDs. Alternatively, both the Functional Reach tests and One Leg stance and variants are limited to only a single motor task, and show very low levels of evidence, hampering their generalizability.

Strengths and limitations of the study

To perform the meta-analyses, data were extracted directly from the included case-control studies. Specific categorization was necessary to reduce the overall noise within this data set. Despite this subgroup analyses, general heterogeneity remained, mostly due to different reporting styles (mean scores, standard deviations, interquartile ranges) and the different level of detail in the descriptions of included participants and methodologies. The larger 95% confidence intervals of both the pooled SMDs and those at the individual study level of the meta-analyses indicate heterogeneity between groups. This could be explained by heterogeneity both between and within the study samples, inducing higher rates of imprecision. Furthermore, some of the reported results are based on post-hoc subclassification of larger groups (*e.g.* age bands of 7-9 years and ≥10 years). While these subclassifications could introduce some levels of bias, they were defined based on the trends

and data found in literature. Having access to the raw data (for example through improved open-science practices) from the actual studies would help improve the quality of the meta-analyses.

A second limitation is that we excluded articles that did not specifically include case-control studies where TD children were compared to children with pathologies. However, this was necessary to assess the known-groups validity. Furthermore, focusing on only one type of study design allowed for more rigorous evaluations of the overall methodological quality and risk of bias in the included articles through the SIGN and GRADE assessment scales. However, with most of the available data expressed as mean performance-based outcomes, it was impossible to determine the means of specific age groups.

An important strength of this review is its clinical approach to the topic of postural control assessment in children. With an overabundance of available pediatric postural control tests, therapists often have a hard time deciding which test is most appropriate for their specific patient. Previous systematic reviews have already addressed this by reporting the psychometric properties of these outcome measures.[11, 12] The fact that the current review builds on these previous studies only strengthens the body of evidence concerning these outcome measures and allows therapists to make a more informed and evidence-based decision when choosing the right performance-based postural control test.

In line with the previous review,[11] our review also only included those pediatric postural control tests with reported normative or reference data. Due to the heterogeneity of our available data and the fact that the available reference data often spans multiple age groups, reference data was not yet included in our analysis. However, future research could aim to investigate how the performances of various types of children (both TD and those with underlying pathologies) compares to these normative datasets. By performing such an analysis, researchers and clinicians alike could better gauge the external validity of this reference data for diagnostic purposes in a clinical context. That way, clinicians would know the usefulness of their available tools, while researchers could identify if certain geographical, cultural, or demographic characteristics influence their applicability to differentiate cases from controls.

Implications for clinical practice and future research

As shown by the examples above, heterogeneity within the dataset was one of the main obstacles for the meta-analyses presented in this review. Future research should thus prioritize sound methodological studies, which aim to evaluate the differences between strictly defined homogenous groups on postural control tests. This review could also act as a first step in constructing and analyzing one large dataset which includes all available raw data to enable analysis of the specific factors and compare these results with the available reference data.

There does not seem to be one postural control test which is able to differentiate unequivocally whether a postural control deficit is present or not, irrespective of age and pathology. However, given the results from our meta-analysis, test batteries – and more specifically the PBS – show the most promise. These recommendations should be made cautiously since only the PBS was implemented most and often in those pathological groups where children showed more severe levels of postural control deficits. Future investigation of the PBS in populations with much milder postural control deficits and other test batteries as defined by Johnson *et al.* (2023) (*e.g.* Kids-BESTest, Fullerton Advanced Balance) in various pediatric populations are in order. More specifically, our findings reaffirm the practical suggestion made by previous authors. When pediatric clinicians merely attempt to identify the presence of a general postural control deficit, larger test batteries of postural control should be favored over single-task tests. When they aim to assess one specific subsystem or underlying function of postural control, single-task tests could still be implemented. However, clinicians should ensure the motor task required for the test they selected is suited for their targeted patient's age and underlying pathology.

Although we tried to establish known-groups validity using cases and controls, none of the included studies compared their groups to normative data nor did they map the degree to which case children actually reported complaints at the time of the study. Future research which addresses these issues could provide new insights into the validity, sensitivity, and specificity of functional postural control tests.

Another aspect this study was unable to account for was the quality of movement during testing. This may be more relevant in milder motor disorders in which the typical developmental milestones are met yet, the difference lies in subtle motor quality deviations. The scores of simple pediatric postural control tests are often purely based on objective quantitative results (*e.g.* time required to complete task, number of repetitions within allotted time, etc.) while often neglecting the role of qualitative performance markers. The PBS is one of the only tests included in this review which incorporates some modicum of qualitative assessment. The RTs on the other hand are only scored

using quantitative data. Future research should therefore investigate how qualitative markers can be implemented in existing functional postural control tests. The inclusion of this aspect of testing might enable the identification of slower abnormal developmental patterns currently missed by clinical tests and thereby refine our understanding of postural control deficits in pediatric populations.

Additionally, technological innovation in both clinical practice and research settings may aid in the pursuit of valid clinical tests.[10, 79] By enabling therapists and researchers to pool the data provided by functional postural control tests, add images or video recordings to test results, and automate specific aspects of testing (*e.g.* timing, measuring), interpretation of test results may become more efficient. Furthermore, reference data could be constantly and easily collected, updated, and sorted according to gender, age, overarching pathology group or even geographical location.

Conclusions

None of the included functional postural control tests exceeded a low level of evidence, which impacts their usefulness in clinical practice for any pediatric population. The type of underlying pathology has a larger impact on the known-groups validity of these tests than age. Thus, it is crucial that therapists select postural control tests based on both the underlying pathology, the motor repertoire and age of their patient and ensure the complexity of the motor tasks, performed in these tests is suited to differentiate performances of their patient and TD peers. Test batteries such as the PBS seem to be more promising than single tests as they cover multiple aspects of postural control. This systematic review and meta-analysis are a first step in determining the clinical applicability of functional postural control tests for which age-specific normative data already exist. Due to the low level of evidence, future research should not only strive to expand and homogenize the available data, but also to improve the levels of available evidence in order to establish the sensitivity and specificity of these tests and their respective reference data. These efforts may ensure that future clinicians are able to select viable and useful functional postural control tests in their daily clinical practice.

References

1. Verbecque E, Johnson C, Rameckers E, Thijs A, van der Veer I, Meyns P, *et al.* Balance control in individuals with developmental coordination disorder: A systematic review and meta-analysis. Gait Posture 2021;83:268–79.

2. Jain PD, Nayak A, Karnad SD. Relationship between trunk muscle strength, reaching ability and balance in children with Down syndrome - A cross-sectional study. Brain Dev 2022;44:95–104.

3. Ghai S, Hakim M, Dannenbaum E, Lamontagne A. Prevalence of Vestibular Dysfunction in Children With Neurological Disabilities: A Systematic Review. Front Neurol 2019;10:1294.

4. De Kegel A, Maes L, Baetens T, Dhooge I, Van Waelvelde H. The influence of a vestibular dysfunction on the motor development of hearing-impaired children. Laryngoscope 2012;122:2837–43.

5. Pavão SL, Barbosa KA, Sato TO, Rocha NA. Functional balance and gross motor function in children with cerebral palsy. Res Dev Disabil 2014;35:2278–83.

6. Jiandani MP, Mhatre BS. Physical therapy diagnosis: how is it different? J Postgrad Med 2018;64:69–72.

7. de Vet HC, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine: A Practical Guide. Cambridge: Cambridge University Press; 2011.

8. Dijkers MP, Murphy SL, Krellman J. Evidence-based practice for rehabilitation professionals: concepts and controversies. Arch Phys Med Rehabil 2012;93(Suppl):S164–76.

9. Horak FB. Clinical measurement of postural control in adults. Phys Ther 1987;67:1881–5.

10. Mancini M, Horak FB. The relevance of clinical balance assessment tools to differentiate balance deficits. Eur J Phys Rehabil Med 2010;46:239–48.

11. Johnson C, Hallemans A, Goetschalckx M, Meyns P, Rameckers E, Klingels K, *et al.* Psychometric properties of functional postural control tests in children: A systematic review. Ann Phys Rehabil Med 2023;66:101729.

12. Verbecque E, Lobo Da Costa PH, Vereeck L, Hallemans A. Psychometric properties of functional balance tests in children: a literature review. Dev Med Child Neurol 2015;57:521–9.

13. Jette DU, Halbert J, Iverson C, Miceli E, Shah P. Use of standardized outcome measures in physical therapist practice: perceptions and applications. Phys Ther 2009;89:125–35.

14. Blomqvist S, Olsson J, Wallin L, Wester A, Rehn B. Adolescents with intellectual disability have reduced postural balance and muscle performance in trunk and lower limbs compared to peers without intellectual disability. Res Dev Disabil 2013;34:198–206.

15. Ockerman J, Coussens M, Bar-On L. Current balance measuremnt practices among Flemish physiotherapists. Gait & Posture 2022;97:137–8.

16. Portney LG. Foundations of Clinical Research: Applications to Evidence-Based Practice. 4e ed. Portney LG, editor. New York, NY: F. A. Davis Company; 2020.

17. Cumberworth VL, Patel NN, Rogers W, Kenyon GS. The maturation of balance in children. J Laryngol Otol 2007;121:449–54.

18. Charpiot A, Tringali S, Ionescu E, Vital-Durand F, Ferber-Viart C. Vestibulo-ocular reflex and balance maturation in healthy children aged from six to twelve years. Audiol Neurootol 2010;15:203–10.

19. Nolan L, Grigorenko A, Thorstensson A. Balance control: sex and age differences in 9- to 16-year-olds. Dev Med Child Neurol 2005;47:449–54.

20. Micarelli A, Viziano A, Augimeri I, Micarelli B, Alessandrini M. Age-related Assessment of Postural Control Development: A Cross-sectional Study in Children and Adolescents. J Mot Behav 2020;52:418–26.

21. Shumway-Cook A, Woollacott MH. Motor Control: translating research into clinical practice. Fifth edition ed. Philadelphia: Wolters Kluwer; 2017.

22. Franjoine MR, Darr N, Held SL, Kott K, Young BL. The performance of children developing typically on the pediatric balance scale. Pediatr Phys Ther 2010;22:350–9.

23. Deitz JC, Richardson P, Atwater SW, Crowe TK, Odiorne M. Performance of Normal Children on the Pediatric Clinical Test of Sensory Interaction for Balance. Occup Ther J Res 1991;11:336–56.

24. Donahoe B, Turner D, Worrell T. The Use of Functional Reach as a

Measurement of Balance in Boys and Girls Without Disabilities Ages 5 to 15 Years. Pediatr Phys Ther 1994;6.

25. Southard V, editor. Development Of Normative Data Using The Mini-BESTest On Children Ages 7-17. 2020.

26. Lissauer T, Carroll W. Illustrated Textbook of Paediatrics. In: Tom Lissauer, Carroll W, editors. Illustrated Textbook of Paediatrics. Sixth Edition. Poland: Elsevier Limited; 2022. p. 52-70.

27. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, *et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71.

28. SIGN. Methodology Checklist 4: Case-control studies. Edinburgh: SIGN; 2012.

29. Schünemann H, Brożek J, Guyatt G, Oxman A. GRADE handbook for grading quality of evidence and strength of recommendations 2013 Updated October 2013.

30. Abdul Rahman RA, Rafi F, Hanapiah FA, Nikmat AW, Ismail NA, Manaf H. Effect of Dual-Task Conditions on Gait Performance during Timed Up and Go Test in Children with Traumatic Brain Injury. Rehabil Res Pract 2018;2018:2071726.

31. Bricout VA, Pace M, Dumortier L, Miganeh S, Mahistre Y, Guinot M. Motor capacities in boys with high functioning autism: which evaluations to choose? J Clin Med 2019;8:1521.

32. Buker N, Salik Sengul Y, Ozbek A. Physical Fitness and Dynamic Balance in Medication Naïve Turkish Children with ADHD. Percept Mot Skills 2020;127:858–73.

33. Dewar RM, Tucker K, Claus AP, Ware RS, Johnston LM. Postural Control Performance on the Functional Reach Test: Validity of the Kids-Balance Evaluation Systems Test (Kids-BESTest) Criteria. Arch Phys Med Rehabil 2021;102:1170–9.

34. Dik J, Saglam M, Tekerlek H, Vardar-Yagli N, Calik-Kutukcu E, Inal-Ince D, *et al.* Visuomotor reaction time and dynamic balance in children with cystic fibrosis and non-cystic fibrosis bronchiectasis: A case-control study. Pediatr Pulmonol 2020;55:2341–7.

35. Jayakaran P, Mitchell L, Johnson GM. Peripheral sensory information and postural control in children with strabismus. Gait Posture 2018;65:197–202.

36. Kalyani HH, Wanigasinghe J. Assessment of the Balance Functions of Children with Down SyndromeSelected Paediatric Clinical Settings in Colombo District, Sri Lanka. Sri Lanka J Child Health 2021;50:239–45.

37. Karakoc K, Mujdeci B. Evaluation of balance in children with sensorineural hearing loss according to age. Am J Otolaryngol 2021;42:102830.

38. Kaya S, Bas B, Er S, Keseroglu K, Korkmaz H. Cervical vestibular-evoked myogenic potentials and balance testing in children with down syndrome. Int Arch Otorhinolaryngol 2021;25:e580–4.

39. Leizerowitz G, Katz-Leurer M. Feasibility, stability and validity of the four square step test in typically developed children and children with brain damage. Brain Inj 2017;31:1356–61.

40. Lukács A, Mayer K, Juhász E, Varga B, Fodor B, Barkai L. Reduced physical fitness in children and adolescents with type 1 diabetes. Pediatr Diabetes 2012;13:432–7.

41. Melo RS, Lemos A, Macky CF, Raposo MC, Ferraz KM. Postural control assessment in students with normal hearing and sensorineural hearing loss. Braz J Otorhinolaryngol 2015;81:431–8.

42. Melo RS, Lemos A, Raposo MC, Belian RB, Ferraz KM. Balance performance of children and adolescents with sensorineural hearing loss: repercussions of hearing loss degrees and etiological factors. Int J Pediatr Otorhinolaryngol 2018;110:16–21.

43. Newman MA, Hirsch MA, Peindl RD, Habet NA, Tsai TJ, Runyon MS, *et al.*; Carolinas Trauma Network Research Group. Reliability of the sub-components of the instrumented timed up and go test in ambulatory children with traumatic brain injury and typically developed controls. Gait Posture 2018;63:248–53.

44. Nicolini-Panisson RD, Donadio MV. Normative values for the Timed 'Up and Go' test in children and adolescents and validation for individuals with Down syndrome. Dev Med Child Neurol 2014;56:490–7.

45. Pace M, Bricout VA. Low heart rate response of children with autism spectrum disorders in comparison to controls during physical exercise. Physiol Behav 2015;141:63–8.

46. Sanz-Santiago V, Diez-Vega I, Donadio MV, Schiwe D, Lopez-Nuevo C, Vendrusculo FM, *et al.* Comparison of physical fitness between healthy and mild-to-moderate asthmatic children with exercise symptoms: A cross-sectional study. Pediatr Pulmonol 2021;56:2512–21.

47. Soares LM, Rozane JM, Carvalho RP. Motor performance of children with cerebral palsy in anterior reach. Clin Biomech (Bristol, Avon) 2019;68:158–62.

48. Alkan H, Mutlu A, Fırat T, Bulut N, Karaduman AA, Yılmaz OT. Effects of functional level on balance in children with Duchenne Muscular Dystrophy. Eur J Paediatr Neurol 2017;21:635–8.

49. Bas B, Karabulut M, Mujdeci B. Evaluation of Balance Test Outcomes in Children with Poor Vision. Cyprus Journal of Medical Sciences. 2020;5:284–8.

50. Coffey C, Sheehan D, Faigenbaum AD, Healy S, Lloyd RS, Kinsella S. Comparison of fitness levels between elementary school children with autism spectrum disorder and age-matched neurotypically developing children. Autism Res 2021;14:2038–46.

51. De Kegel A, Dhooge I, Cambier D, Baetens T, Palmans T, Van Waelvelde H. Test-retest reliability of the assessment of postural stability in typically developing children and in hearing impaired children. Gait Posture 2011;33:679–85.

52. De Kegel A, Dhooge I, Peersman W, Rijckaert J, Baetens T, Cambier D, *et al.* Construct validity of the assessment of balance in children who are developing typically and in children with hearing impairments. Phys Ther 2010;90:1783–94.

53. Deforche BI, Hills AP, Worringham CJ, Davies PS, Murphy AJ, Bouckaert JJ, *et al.* Balance and postural skills in normal-weight and overweight prepubertal boys. Int J Pediatr Obes 2009;4:175–82.

54. Franjoine MR, Darr N, Young B, McCoy SW, LaForme Fiss A. Examination of the effects of age, sex, and motor ability level on balance capabilities in children with cerebral palsy GMFCS levels I, II, III and typical development using the Pediatric Balance Scale. Dev Neurorehabil 2022;25:115–24.

55. Goldman S, McCullough AK, Young SD, Mueller C, Stahl A, Zoeller A, *et al.* Quantitative gait assessment in children with 16p11.2 syndrome. J Neurodev Disord 2019;11:26.

56. Katz-Leurer M, Rotem H, Keren O, Meyer S. Balance abilities and gait characteristics in post-traumatic brain injury, cerebral palsy and typically developed children. Dev Neurorehabil 2009;12:100–5.

57. Katz-Leurer M, Rotem H, Keren O, Meyer S. Recreational physical activities among children with a history of severe traumatic brain injury. Brain Inj 2010;24:1561–7.

58. Katz-Leurer M, Rotem H, Lewitus H, Keren O, Meyer S. Functional balance tests for children with traumatic brain injury: within-session reliability. Pediatr Phys Ther 2008;20:254–8.

59. Kembhavi G, Darrah J, Magill-Evans J, Loomis J. Using the berg balance scale to distinguish balance abilities in children with cerebral palsy. Pediatr Phys Ther 2002;14:92–9.

60. Marchese V, Rock K, York T, Creath R, Gray V. Neuromuscular mechanisms that contribute to gross motor performance in survivors of childhood acute lymphoblastic leukemia. J Pediatr Rehabil Med 2021;14:415–23.

61. Mohammed AH. Correlation between Balance and Attention in Children with Attention Deficit Hyperactivity Disorder. International Journal of Pharmaceutical Research and Allied Sciences. 2019;8:126–30.

62. Newman MA, Hirsch MA, Peindl RD, Habet NA, Tsai TJ, Runyon MS, *et al.*; Carolinas Trauma Network Research Group. Use of an instrumented dual-task timed up and go test in children with traumatic brain injury. Gait Posture 2020;76:193–7.

63. Promsorn S, Taweetanalarp S. The multi-directional reach test in children with Down syndrome. Hong Kong Physiother J 2021;41:65–74.

64. Tsiros MD, Buckley JD, Howe PR, Olds T, Walkley J, Taylor L, *et al.* Day-to-day physical functioning and disability in obese 10- to 13-year-olds. Pediatr Obes 2013;8:31–41.

65. Zaino CA, Marchese VG, Westcott SL. Timed up and down stairs test: preliminary reliability and validity of a new measure of functional mobility. Pediatr Phys Ther 2004;16:90–8.

66. Roostaei M, Raji P, Kalantari KK, Faghihzadeh E, Fragala-Pinkham M. Effect of upper extremity constraints on functional and dynamic postural control in children with hemiplegic cerebral palsy. Dev Neurorehabil 2022;25:281–8.

67. Walowska J, Bolach E. Evaluation of general physical fitness in hard of hearing and hearing children. Fizjoterapia. 2011;19:9–27.

68. Graff K, Szczerbik E, Kalinowska M, Kaczmarczyk K, Stępień A, Syczewska M. Using the TUG Test for the Functional Assessment of Patients with Selected Disorders. Int J Environ Res Public Health 2022;19:4602.

69. Haibach PS, Lieberman LJ. Balance and self-efficacy of balance in children with CHARGE syndrome. J Vis Impair Blind 2013;107:297–309.

70. Huang PC, Pan PJ, Ou YC, Yu YC, Tsai YS. Motion analysis of throwing Boccia balls in children with cerebral palsy. Res Dev Disabil 2014;35:393–9.

71. Henderson SE, Sugden DA, Barnett AL. Movement Assessment Battery for Children-2 (Mabc-2). In: Sugden DA, editor. 2nd ed. London, UK.: The Psychological Corporation; 2007.

72. Bruininks RH, Bruininks BD. Bruininks-Oseretsky Test of Motor Proficiency: Manual. In: Bruininks BD, editor. 2nd ed. Minneapolis, MN, USA: Pearson Assessments; 2005.

73. Deshmukh AA, Ganesan S, Tedla JS. Normal values of functional reach and lateral reach tests in Indian school children. Pediatr Phys Ther 2011;23:23–30.

74. Yuksel E, Ozcan Kahraman B, Nalbant A, Kocak UZ, Unver B. Functional Reach and Lateral Reach Tests in Turkish Children. Phys Occup Ther Pediatr 2017;37:389–98.

75. Deshmukh AA. Normal values of functional reach and lateral reach tests in children with knee hypermobility. Pediatr Phys Ther 2014;26:230–6.

76. Sharma K, Samuel AJ, Midha D, Aranha VP, Narkeesh K, Arumugam N. Multi-directional reach test in South Asian children: normative reference scores from 5 year to 12 years old. Homo 2018;69:62–9.

77. Habib Z, Westcott SL. Assessment of Anthropometric Factors on Balance Tests in Children. Pediatr Phys Ther 1998;10:101–9.

78. Condon C, Cremin K. Static balance norms in children. Physiother Res Int 2014;19:1–7.

79. Bonnechère B, Timmermans A, Michiels S. Current Technology Developments Can Improve the Quality of Research and Level of Evidence for Rehabilitation Interventions: A Narrative Review. Sensors (Basel) 2023;23:875.

Conflicts of interest
The authors certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript.

Authors' contributions
Jorn Ockerman and Evi Verbecque have given substantial contributions to the conception or the design of the manuscript and have participated to drafting the manuscript; Jorn Ockerman, Evi Verbecque and Silke Velghe have given substantial contributions to acquisition, analysis, and interpretation of the data; Evi Lynn Bar-On revised the manuscript critically. Evi Verbecque and Lynn Bar-On share last authorship. All authors read and approved the final version of the manuscript.

Funding
This research was funded by a Baekeland grant from VLAIO (HBC.2020.2294)

History
Article first published online: June 5, 2024. - Manuscript accepted: May 8, 2024. - Manuscript revised: March 14, 2024. - Manuscript received: August 9, 2023.

SUPPLEMENTARY DIGITAL MATERIAL 1

Search Strategy

Pubmed

("Balance Error Scoring System"[Title/abstract] OR "Tandem Stance"[Title/abstract] OR "Flamingo test"[Title/abstract] OR "Stork Balance Stand Test"[Title/abstract] OR "Star Excursion Balance Test"[Title/abstract] OR "Pediatric Reach Test"[Title/abstract] OR "Functional Reach Test"[Title/abstract] OR "Lateral Reach Test"[Title/abstract] OR "Multi-directional Reach Test"[Title/abstract] OR "Timed Up and Go test"[Title/abstract] OR "Timed Get up and go test"[Title/abstract] OR "Get up and go test"[Title/abstract] OR "Complex Gait Test"[Title/abstract] OR "Balance Beam Walking"[Title/abstract] OR "Pediatric Balance Scale"[Title/abstract] OR "Berg Balance Scale"[Title/abstract] OR "Ghent Developmental Balance Test"[Title/abstract] OR PREFIT[Title/abstract] OR Eurofit[Title/abstract] OR PERFFIT[Title/abstract]) AND (child[MeSH Terms] OR child[Title/Abstract] OR children[title/abstract] OR "preschool child"[Title/abstract] OR preschooler[Title/abstract] OR youth[Title/abstract] OR minor[Title/abstract] OR adolescent[Title/abstract] OR adolescents[Title/abstract])

Web of Science

("Balance Error Scoring System" OR "Tandem Stance" OR "Flamingo test" OR "Stork Balance Stand Test" OR "Star Excursion Balance Test" OR "Pediatric Reach Test" OR "Functional Reach Test" OR "Lateral Reach Test" OR "Multi-directional Reach Test" OR "Timed Up and Go test" OR "Timed Get up and go test" OR "Get up and go test" OR "Complex Gait Test" OR "Balance Beam Walking" OR "Pediatric Balance Scale" OR "Berg Balance Scale" OR "Ghent Developmental Balance Test" OR PREFIT OR Eurofit OR PERFFIT) AND (child OR child OR children OR "preschool child" OR preschooler OR youth OR minor OR adolescent OR adolescents)

Scopus

(TITLE-ABS-KEY("Balance Error Scoring System") OR TITLE-ABS-KEY("Tandem Stance") OR TITLE-ABS- KEY("Flamingo test") OR TITLE-ABS-KEY("Stork Balance Stand Test") OR TITLE-ABS-KEY("Star Excursion Balance Test") OR TITLE-ABS-KEY("Pediatric Reach Test") OR TITLE-ABS-KEY("Functional Reach Test") OR TITLE-ABS-KEY("Lateral Reach Test") OR TITLE-ABS-KEY("Multi-directional Reach Test") OR TITLE- ABS-KEY("Timed Up and Go test") OR TITLE-ABS-KEY("Timed Get up and go test") OR TITLE-ABS- KEY("Get up and go test") OR TITLE-ABS-KEY("Complex Gait Test") OR TITLE-ABS-KEY("Balance Beam Walking") OR TITLE-ABS-KEY("Pediatric Balance Scale") OR TITLE-ABS-KEY("Berg Balance Scale") OR TITLE-ABS-KEY("Ghent Developmental Balance Test") OR TITLE-ABS-KEY("PREFIT") OR TITLE-ABS- KEY("Eurofit") OR TITLE-ABS-KEY("PERFFIT")) AND (INDEXTERMS("child") OR TITLE-ABS-KEY("child") OR TITLE-ABS-KEY("children") OR TITLE-ABS-KEY("preschool child") OR TITLE-ABS-KEY("preschooler") OR TITLE-ABS-KEY("youth") OR TITLE-ABS-KEY("minor") OR TITLE-ABS-KEY("adolescent") OR TITLE-ABS-KEY("adolescents")

SUPPLEMENTARY DIGITAL MATERIAL 2

Supplementary Table I.—Description of the cases and controls participating in the original studies.

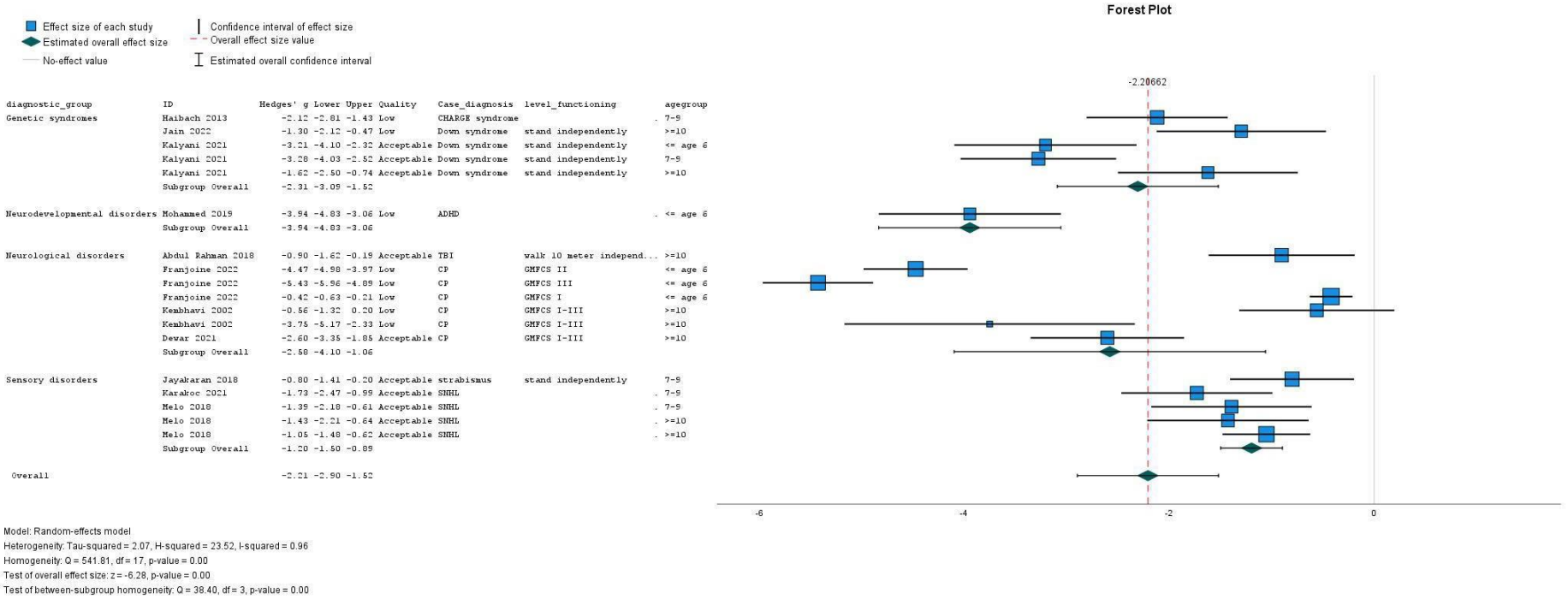
AUTHOR	CASES					CONTROLS				FUNTIONAL TEST USED
	Pathology-specific diagnosis	Age (years) (mean, SD, range)	Sex (m/f)	Height (cm) (mean, SD)	Weight (kg) (mean, SD)	Age (years) (mean, SD, range)	Sex (m/f)	Height (cm) (mean, SD)	Weight (kg) (mean, SD)	
Abdul Rahman 2018 [30]	TBI	11.6 (2), -	12/2	141 (17)	40.4 (12.5)	11.4 (2.3)	17/4	142 (15)	41.2 (17.1)	PBS (score)
Alkan 2017 [48]	DMD	8.8 (1.4), -	52/0	123.3 (9.6)	28.1 (6.9)	8.6 (1.6)	17/0	126.4 (11.7)	30.4 (9.8)	expanded TUG (s)
Bas 2020 [49]	congenital poor vision	7.4 (1.9)	n=20			8.2 (1.1)	n=20			FRT (cm), TUG (s)
Bricout 2019 [31]	ASD	10.7 (1.3), 8-12	22/0	144.7 (8.7)	36.0 (13.3)	10.0 (1.6), 8-12	20/0	141.0 (10.5)	33.3 (7.2)	Flamingo balance (s), sustained postures (tandem, OLS, s), tandem walk backward & forward (errors)
Buker 2020 [32]	ADHD	9 (2.75), 7-12	20/4	140.0 (13.0)	37.3 (9.6)	9.0 (3.0), 7-12	15/4	133.0 (16.3)	32.4 (17.4)	Flamingo balance (number of falls), Y-balance (cm)
Coffey 2021 [50]	ASD	8.3 (2.1), 4-13	77/15	135.4 (14.1)	37.0 (14.9)	8.3 (2.1), 4-13	70/82	135.9 (15.1)	32.9 (11.6)	Stork balance test (s)
Deforche 2009 [53]	overweight	9.3 (1.0), 8-10	25/0	140.7 (6.4)	47.2 (7.9)	9.3 (0.8), 8-10	25/0	136.2 (6.8)	30.4 (4.3)	heel-to-toe walk (n steps), OLS on a beam (non-)dominant leg (s, max 10)
De Kegel 2010 [52]**, 2011** [51]	hearing impairment	9.4 (1.9), 6-12	13/10			9.6 (2.0), 6-12	27/22			OLS EC (s), OLS EC (s)
Dewar 2021 [33]	CP	11.7 (2.7), 7-16	11/6	146.1 (14.4)	41.1 (15.8)	10.9 (2.3), 7-17	21/20	150.0 (13.9)	38.9 (14.6)	Kids-BESTest
Dik 2020 [34]	cystic fibrosis	10.7 (2.9), 6-18	7/7			11.4 (3.3), 6-18	10/11			PRT (cm)
Franjoine 2022 [54]	CP GMFCS I	-, range: 2-4	76/63			-, 2-4	144/114			PBS (score)
	CP, GMFCS II	-, range: 2-4	19/16			-, 2-4	144/114			
	CP, GMFCS III	-, range: 2-4	24/21			-, 2-4	144/114			
Goldman 2019 [55]	Genetic variant 16p11.2	9.3 (5.4), -	10/11			7.6 (7.8), -	9/6			TUG (s)
Haibach 2013 [69]	CHARGE syndrome	8.5 (2.09), -	13/9			9.3 (1.8), -	31			PBS (score)
Huang 2014 [70]	CP	10.1 (2.1), -	6/7	138.6 (11.2)	32.2 (11.7)	9.0 (1.6), -	11/9	135.6 (9.8)	32.2 (10.4)	Pediatric reach test (cm)

Jain 2022 [2]	Down syndrome	10.7 (3), -	6/8	136.11 (16.63)	32.93(11.84)	10.71 (3), -	6/8	137.93 (16.04)	37.75 (10.87)	Modified FRT, PBS
Jayakaran 2018 [35]	SNHL	7.2 (1.8); 5-10	14/7	120 (10)	26.2 (6.7)	7.6 (1.8), -	12/13	130 (10)	26.6 (6.4)	PBS
Kalyani 2021 [36]	Down syndrome	-, range: 3-12	64			-, range: 3-12	80			PBS
Karakoc 2021 [37]	SNHL	8.1 (1.1), 6-10	12/8			-, range: 6-10	11/9			FRT (cm), PBS, OLS (s), TUG (s)
Kaya 2021 [38]	Down syndrome	10.2 (0.67), 9-11	8/7			10.2 (0.67), 9-11	8/7			PBS, TUG
Katz Leurer 2008 [58]	TBI	8.7 (3.5), -	16/8	130 (22)	30.5 (15)	8.5 (3.0), -	16/8	135 (18)	31 (11)	FRT (cm), TUG (s)
Katz Leurer 2009 [56]	TBI	9 (2), -	10/5	126 (14)	25 (9)	9 (2), -	18/12	130 (17)	29 (11)	
	CP	9 (2), -	10/5	124 (14)	24 (8)	9 (2), -	18/12	130 (17)	29 (11)	
Katz Leurer 2010 [57]	TBI	9.5 (2.2), -	9/6	136 (14)	32.6 (5.8)	9.9 (1.3), -	10/5	142 (10)	32 (9.9)	
Kembhavi 2002 [59]	CP- spastic hemiplegia	10.4(1.5), 8.2-13.0	6/8			10.6 (1.4), 8.1-12.9	11/3			BBS
	CP - spastic diplegia	10.7 (1.2), 8.2-11.9	9/3			10.6 (1.4), 8.1-12.9	11/3			BBS
Leizerowitz 2015 [39]	ABI	9.5 (6.5- 12.9), -	7/5	140 (110-160)	33.5 (17.5-41)	8.6 (6-12.1), -	15/15	1.4 (1.2 - 1.5)	29.3 (18-47)	FSST, TUG (s)
	CP	7.9 (6.7-11.9), -	10/10	120 (140 - 110)	24 (49-16)	8.6 (6-12.1), -	15/15	1.4 (1.2 - 1.5)	29.3 (18-47)	
Lukacs 2012 [40]	Diabetes type I (8-12)	10.6 (1.5), -	25/27			10.9 (1.3), -	28/32		18,12	Flamingo balance (n of error)
Marchese 2021 [60]	Leukemia survivors	11.52 (3.27). 8.7-16.5	3/3	141.5 (19.45)	46.68 (26.43)	10.9 (3.6), 7.1 - 16.1	3/3	145.67 (17.42)	40.83 (13.07)	TUG (s)
Melo 2015 [41]	SNHL	12.5 (3.5); 7-18	24/24			12.5 (3.5); 7-18	24/24			BESS
		-, range: 7-10	n=16			-,range: 7-10	n=16			
		-, range: 11-14	n=16			-,range: 11-14	n=16			
Melo 2018 [42]	SNHL	12.5 (3.5); 7-18	24/24			12.5 (3.5); 7-18	24/24			PBS
		-, range: 7-10	n=16			-, range: 7-10	n=16			
		-, range: 11-14	N=16			-, range:7-10	N=16			
Mohammed 2019 [61]	ADHD	5.7 (0.4), -	22/8			5.6 (0.1), -	20/10			PBS
Newman 2018 [43], **2020** [62]	TBI	10.5 (1.5), -	6/6	148 (15.4)	40.9 (10.3)	10.4 (1.3), -	5/5	147.9 (13.8)	41.1 (8.7)	TUG + DT
Nicolini-Panison 2014 [44]	Down syndrome	10.6 (4.4)	16/24			10.8 (4.4)	459			
Pace & Bricout 2015 [45]	ASD	10.1 (1.4), -	10	140 (10)	32.1 (8.3)	10.2 (1.4), -	10	140 (10)	34.8 (10.2)	Flamingo balance test
Promsorn & Taweetanalarp 2021 [63]	Down syndrome	8.3 (0.6), -	4/11	121.37 (6.72)	25.87 (8.22)	8.3 (0.6), -	4/11	128.61 (4.89)	28.38 (5.43)	MDRT (Backward, forward, leftward, rightward)
		10.8 (1.1), -	7/8	129.6 (6.76)	34.13 (9.86)	10.7 (1.0), -	7/8	143.93 (8.03)	35.55 (5.71)	

Roostaei 2022 [66]	Hemiplegic CP	8.16 (2.28)	7/5	129 (15.09)	28.62 (9.46)	8.25 (2.26)	7/5	133 (14.10)	30.75 (8.84)	TUG, TUG Dual task
Sanz-Santiago 2021 [46]	Asthma	11.5 (2.7), -	71	146.7 (16.2)	43.8 (14.4)	10.7 (2.5), -	71	144.9 (14.1)	39.2 (10.8)	TUDS, TUG
Soares 2019 [47]	Unilateral CP	10.3 (3.3), -	4/2	135 (16)	35.50 (8.96)	9.8 (3.1), -	7/7	140 (19)	36.75 (16.06)	FRT
	Bilateral CP	9.6 (2.5), -	3/5	136 (13)	32.66 (12.37)	9.8 (3.1), -	7/7	140 (19)	36.75 (16.06)	
Tsiros 2012 [64]	Obese	11.8 (0.1), -	56/51	155 (1)	72 (1.4)	12.0 (0.1), -	76/56	154 (1)	43.7 (0.7)	TUG
Zaino 2004 [65]	CP (GMFCS I)	11.3 (1.4), 8.1 -14.9	0/4	139.9 (5.3), 131.0-155.0	37.8 (7.9), 28.2-61.4	10.4 (0.4), 8.0-14.0	0/14	142 (2.9), 124.5-160.0	37 (2.7), 22.7-53.2	TUG (s), TUDS (s), OLS (s), FRT (cm)
		10.4 (1.1), 8.3-14.4	5/0	143.5 (7.9), 123.5-167	38.8 (4.7), 28.2 - 52.3	11.9 (0.8), 8.1-14.11	13/0	154(5.8), 119.5-187.5	45.7 (5.4), 19.1-82.7	
	CP (GMFCS II/III)	10.4 (1.1), 8.0 - 13.7	0/5	140 (5.6), 123.5-153.0	40.4 (11), 22.7-82.3	10.4 (0.4), 8.0-14.0	0/14	142 (2.9), 124.5-160.0	37 (2.7), 22.7-53.2	
		11.4 (0.8), 8.1 -13.6	6/0	149 (6.4), 119.5-166.0	38.7 (5.5), 14.5-51.4	11.9 (0.8), 8.1-14.11	13/0	154(5.8), 119.5-187.5	45.7 (5.4), 19.1-82.7	

SUPPLEMENTARY DIGITAL MATERIAL 3

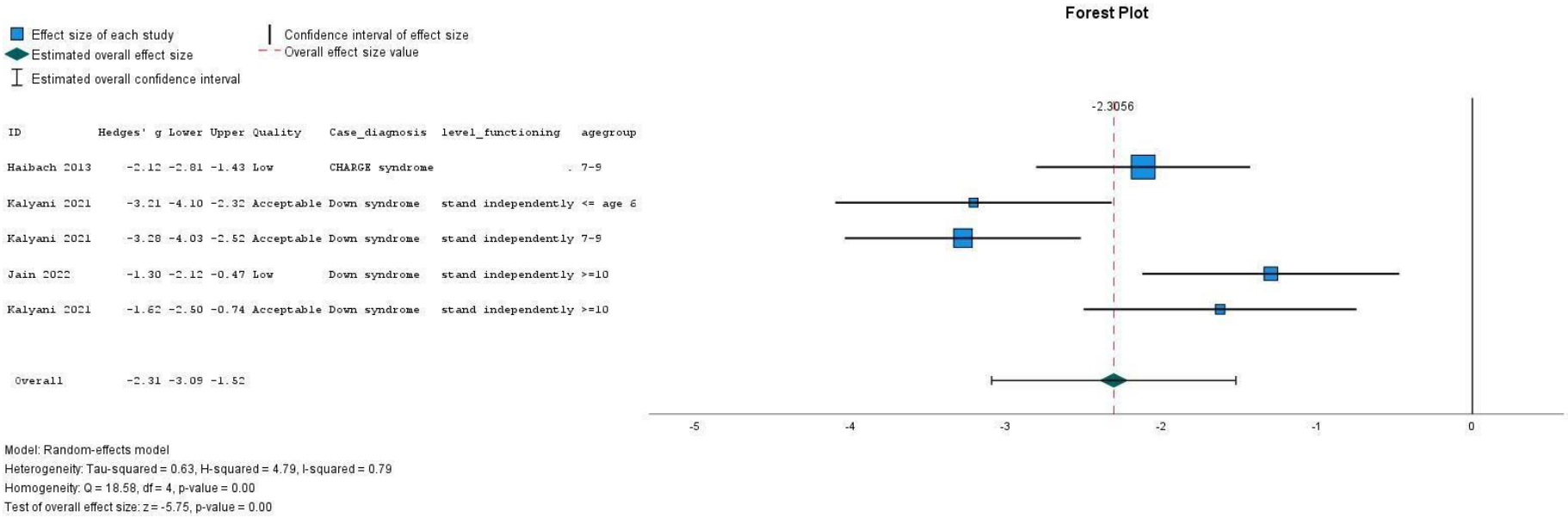
Supplementary Figure 1.—Differences in performances on test batteries between cases and controls.



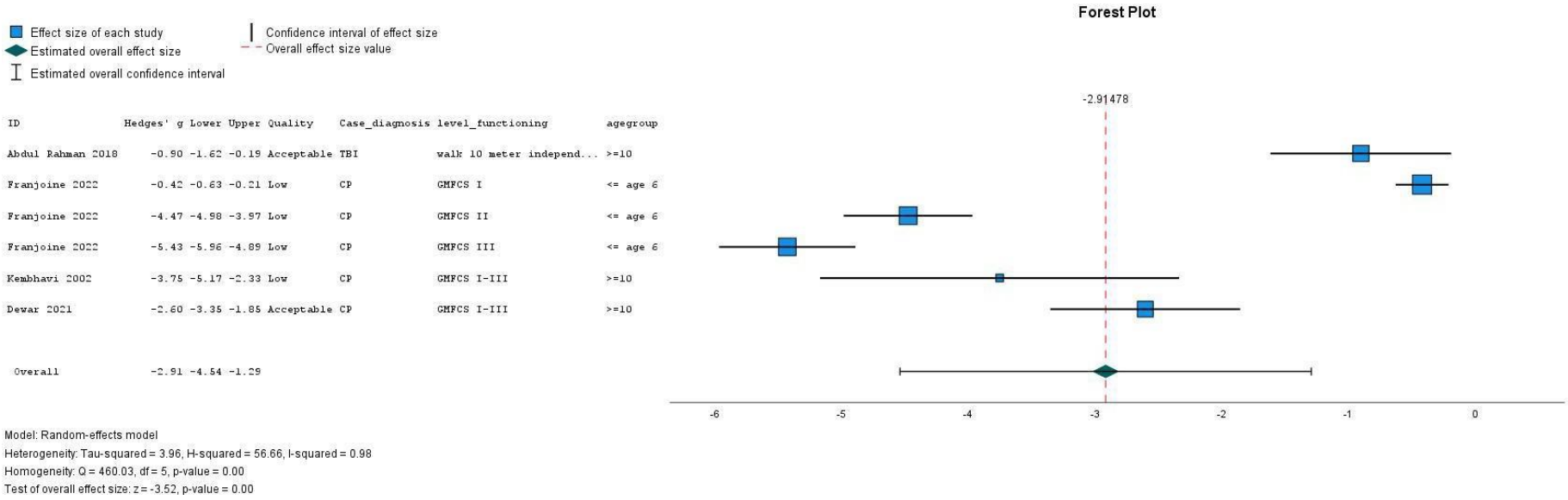
diagnostic_group	ID	Hedges' g	Lower	Upper	Quality	Case_diagnosis	level_functioning	agegroup
Genetic syndromes	Haibach 2013	-2.12	-2.81	-1.43	Low	CHARGE syndrome		7-9
	Jain 2022	-1.30	-2.12	-0.47	Low	Down syndrome	stand independently	>=10
	Kalyani 2021	-3.21	-4.10	-2.32	Acceptable	Down syndrome	stand independently	<= age 6
	Kalyani 2021	-3.28	-4.03	-2.52	Acceptable	Down syndrome	stand independently	7-9
	Kalyani 2021	-1.62	-2.50	-0.74	Acceptable	Down syndrome	stand independently	>=10
	Subgroup Overall	-2.31	-3.09	-1.52				
Neurodevelopmental disorders	Mohammed 2019	-3.94	-4.83	-3.06	Low	ADHD		<= age 6
	Subgroup Overall	-3.94	-4.83	-3.06				
Neurological disorders	Abdul Rahman 2018	-0.90	-1.62	-0.19	Acceptable	TBI	walk 10 meter independ...	>=10
	Franjoine 2022	-4.47	-4.98	-3.97	Low	CP	GMFCS II	<= age 6
	Franjoine 2022	-5.43	-5.96	-4.89	Low	CP	GMFCS III	<= age 6
	Franjoine 2022	-0.42	-0.63	-0.21	Low	CP	GMFCS I	<= age 6
	Kembhavi 2002	-0.56	-1.32	0.20	Low	CP	GMFCS I-III	>=10
	Kembhavi 2002	-3.75	-5.17	-2.33	Low	CP	GMFCS I-III	>=10
	Dewar 2021	-2.60	-3.35	-1.85	Acceptable	CP	GMFCS I-III	>=10
	Subgroup Overall	-2.58	-4.10	-1.06				
Sensory disorders	Jayakaran 2018	-0.80	-1.41	-0.20	Acceptable	strabismus	stand independently	7-9
	Karakoc 2021	-1.73	-2.47	-0.99	Acceptable	SNHL		7-9
	Melo 2018	-1.39	-2.18	-0.61	Acceptable	SNHL		7-9
	Melo 2018	-1.43	-2.21	-0.64	Acceptable	SNHL		>=10
	Melo 2018	-1.05	-1.48	-0.62	Acceptable	SNHL		>=10
	Subgroup Overall	-1.20	-1.50	-0.89				
Overall		-2.21	-2.90	-1.52				

Model: Random-effects model
Heterogeneity: Tau-squared = 2.07, H-squared = 23.52, I-squared = 0.96
Homogeneity: Q = 541.81, df = 17, p-value = 0.00
Test of overall effect size: z = -6.28, p-value = 0.00
Test of between-subgroup homogeneity: Q = 38.40, df = 3, p-value = 0.00

Supplementary Figure 2.—Subgroup analysis - Differences in performance on *test batteries* between children with *genetic syndromes* and controls.



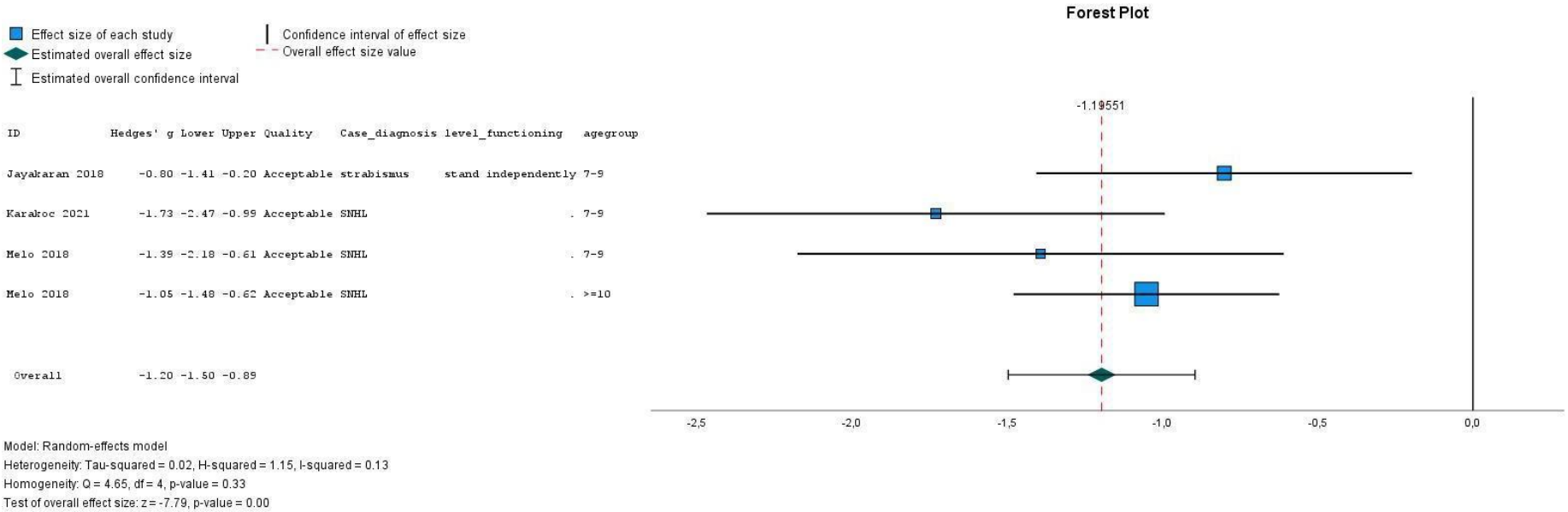
Supplementary Figure 3.—Differences in performance on *test batteries* between *children with neurological disorders* and controls.



ID	Hedges' g	Lower	Upper	Quality	Case_diagnosis	level_functioning	agegroup
Abdul Rahman 2018	-0.90	-1.62	-0.19	Acceptable	TBI	walk 10 meter independ...	>=10
Franjoine 2022	-0.42	-0.63	-0.21	Low	CP	GMFCS I	<= age 6
Franjoine 2022	-4.47	-4.98	-3.97	Low	CP	GMFCS II	<= age 6
Franjoine 2022	-5.43	-5.96	-4.89	Low	CP	GMFCS III	<= age 6
Kembhavi 2002	-3.75	-5.17	-2.33	Low	CP	GMFCS I-III	>=10
Dewar 2021	-2.60	-3.35	-1.85	Acceptable	CP	GMFCS I-III	>=10
Overall	-2.91	-4.54	-1.29				

Model: Random-effects model
Heterogeneity: Tau-squared = 3.96, H-squared = 56.66, I-squared = 0.98
Homogeneity: Q = 460.03, df = 5, p-value = 0.00
Test of overall effect size: z = -3.52, p-value = 0.00

Supplementary Figure 4.—Subgroup analysis - Differences in performance on *test batteries* between *children with sensory deficits* and controls.

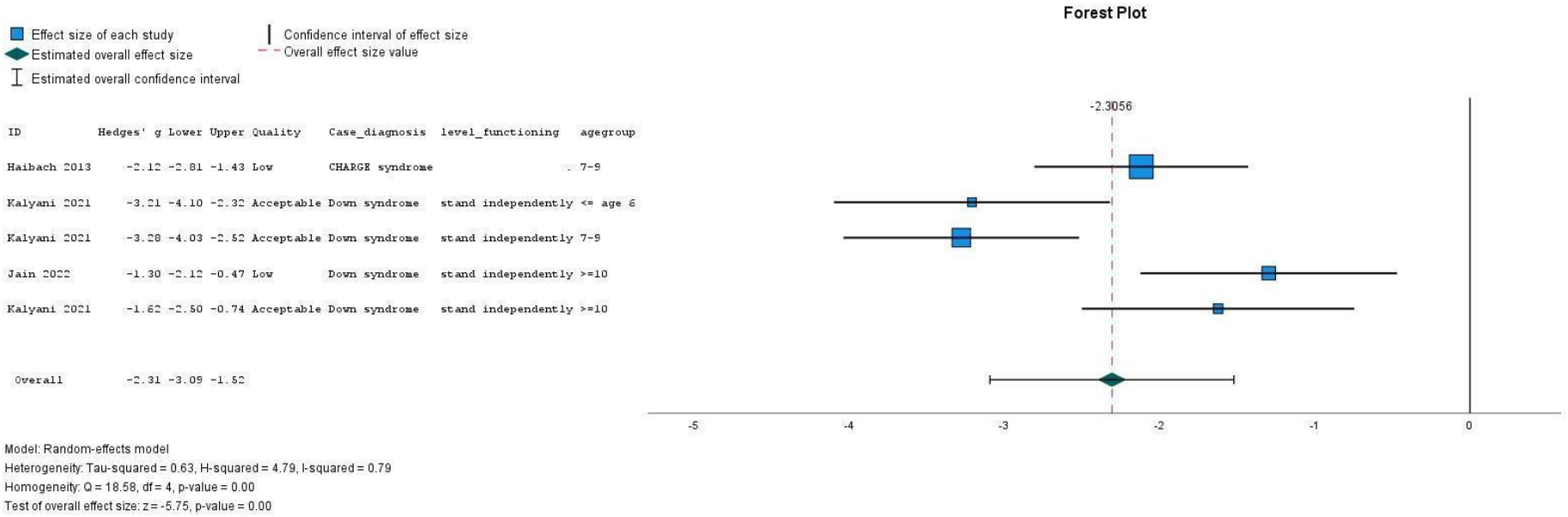


ID	Hedges' g	Lower	Upper	Quality	Case_diagnosis	level_functioning	agegroup
Jayakaran 2018	-0.80	-1.41	-0.20	Acceptable	strabismus	stand independently	7-9
Karakoc 2021	-1.73	-2.47	-0.99	Acceptable	SNHL	.	7-9
Melo 2018	-1.39	-2.18	-0.61	Acceptable	SNHL	.	7-9
Melo 2018	-1.05	-1.48	-0.62	Acceptable	SNHL	.	>=10
Overall	-1.20	-1.50	-0.89				

Model: Random-effects model
Heterogeneity: Tau-squared = 0.02, H-squared = 1.15, I-squared = 0.13
Homogeneity: Q = 4.65, df = 4, p-value = 0.33
Test of overall effect size: z = -7.79, p-value = 0.00

SUPPLEMENTARY DIGITAL MATERIAL 4

Supplementary Figure 2.—Subgroup analysis - Differences in performance on *test batteries* between children with *genetic syndromes* and controls.

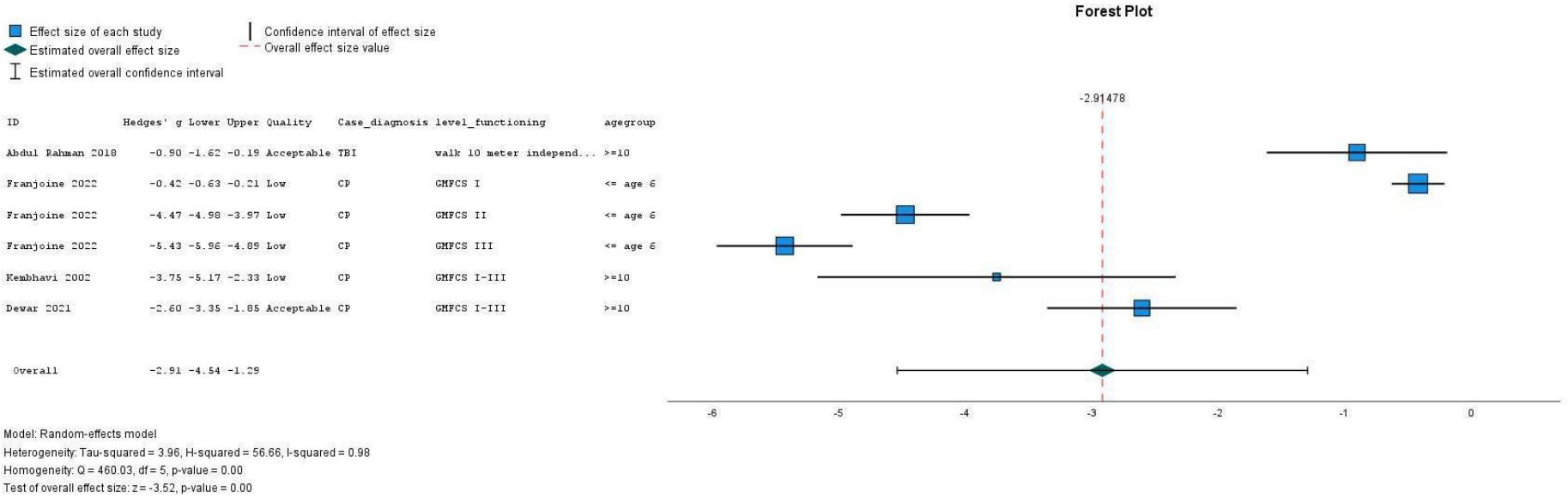


ID	Hedges' g	Lower	Upper	Quality	Case_diagnosis	level_functioning	agegroup
Haibach 2013	-2.12	-2.81	-1.43	Low	CHARGE syndrome	.	7-9
Kalyani 2021	-3.21	-4.10	-2.32	Acceptable	Down syndrome	stand independently	<= age 6
Kalyani 2021	-3.28	-4.03	-2.52	Acceptable	Down syndrome	stand independently	7-9
Jain 2022	-1.30	-2.12	-0.47	Low	Down syndrome	stand independently	>=10
Kalyani 2021	-1.62	-2.50	-0.74	Acceptable	Down syndrome	stand independently	>=10
Overall	-2.31	-3.09	-1.52				

Model: Random-effects model
Heterogeneity: Tau-squared = 0.63, H-squared = 4.79, I-squared = 0.79
Homogeneity: Q = 18.58, df = 4, p-value = 0.00
Test of overall effect size: z = -5.75, p-value = 0.00

SUPPLEMENTARY DIGITAL MATERIAL 5

Supplementary Figure 3.—Differences in performance on *test batteries* between *children with neurological disorders* and controls.

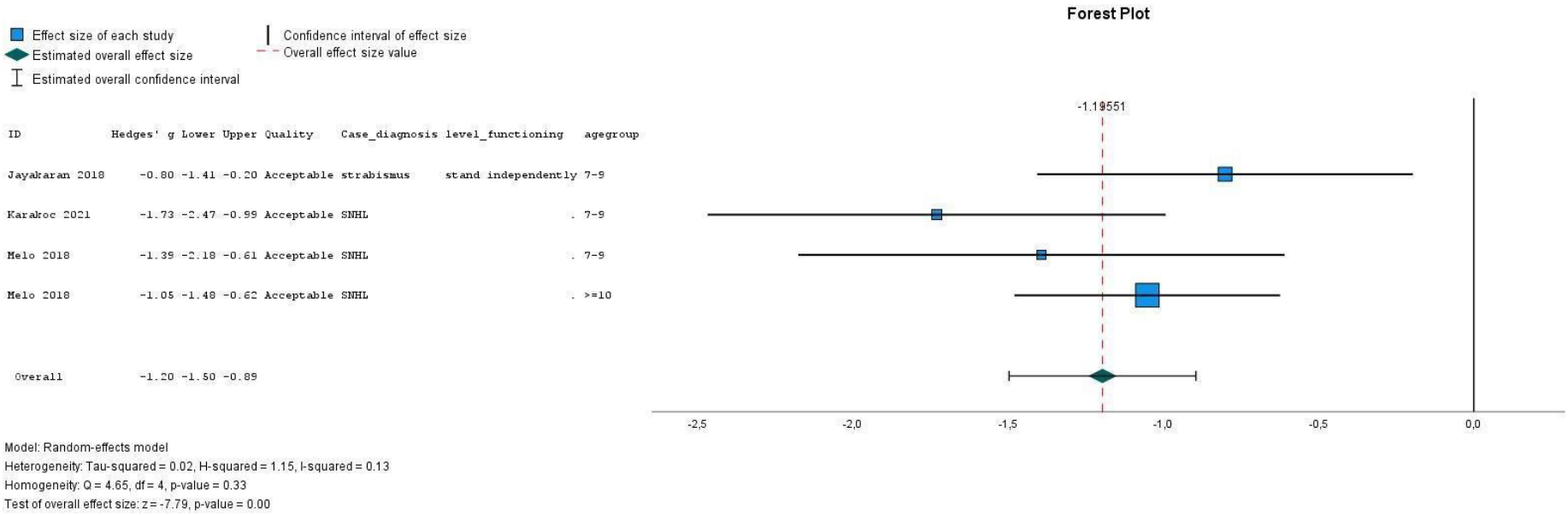


ID	Hedges' g	Lower	Upper	Quality	Case_diagnosis	level_functioning	agegroup
Abdul Rahman 2018	-0.90	-1.62	-0.19	Acceptable	TBI	walk 10 meter independ...	>=10
Franjoine 2022	-0.42	-0.63	-0.21	Low	CP	GMFCS I	<= age 6
Franjoine 2022	-4.47	-4.98	-3.97	Low	CP	GMFCS II	<= age 6
Franjoine 2022	-5.43	-5.96	-4.89	Low	CP	GMFCS III	<= age 6
Kembhavi 2002	-3.75	-5.17	-2.33	Low	CP	GMFCS I-III	>=10
Dewar 2021	-2.60	-3.35	-1.85	Acceptable	CP	GMFCS I-III	>=10
Overall	-2.91	-4.54	-1.29				

Model: Random-effects model
Heterogeneity: Tau-squared = 3.96, H-squared = 56.66, I-squared = 0.98
Homogeneity: Q = 460.03, df = 5, p-value = 0.00
Test of overall effect size: z = -3.52, p-value = 0.00

SUPPLEMENTARY DIGITAL MATERIAL 6

Supplementary Figure 4.—Subgroup analysis - Differences in performance.

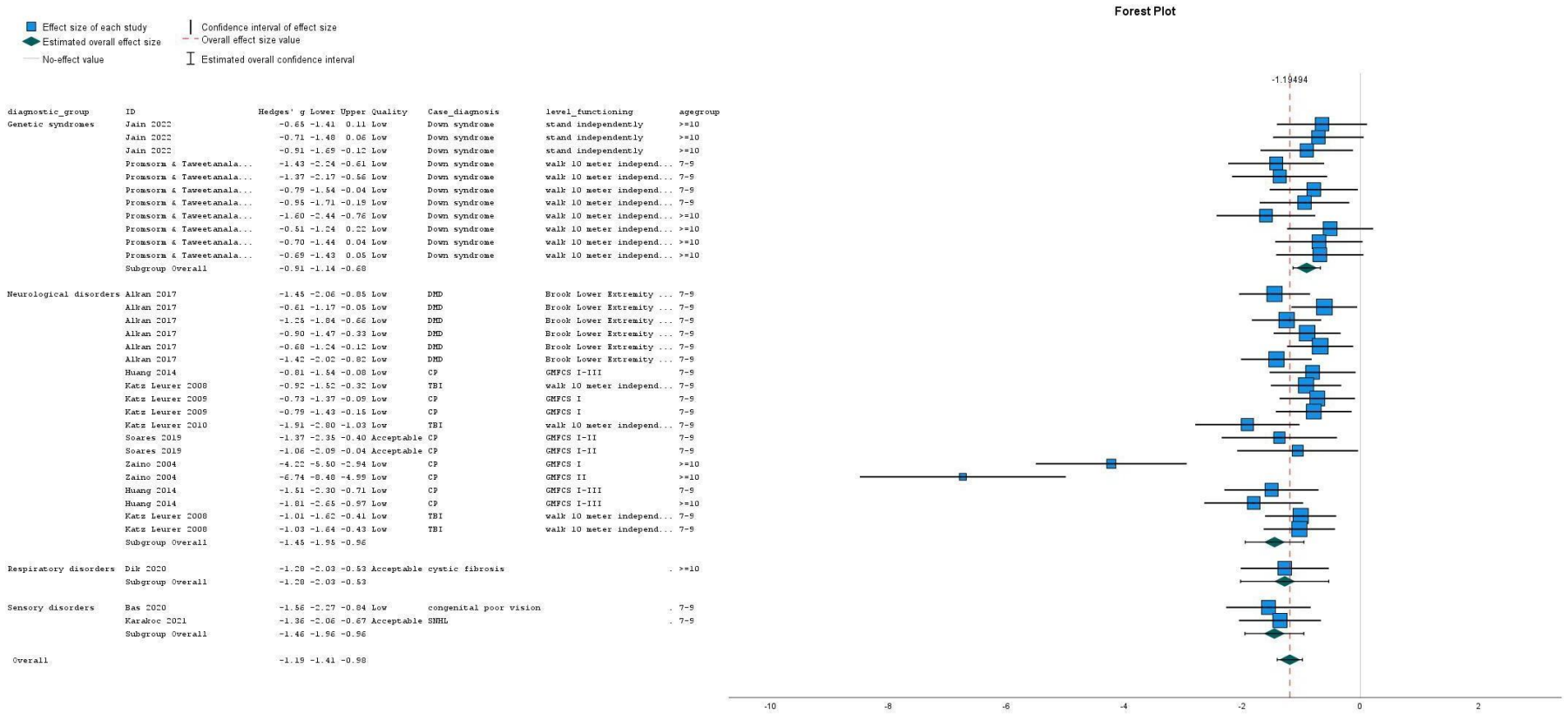


ID	Hedges' g	Lower	Upper	Quality	Case_diagnosis	level_functioning	agegroup
Jayakaran 2018	-0.80	-1.41	-0.20	Acceptable	strabismus	stand independently	7-9
Karakoc 2021	-1.73	-2.47	-0.99	Acceptable	SNHL	.	7-9
Melo 2018	-1.39	-2.18	-0.61	Acceptable	SNHL	.	7-9
Melo 2018	-1.05	-1.48	-0.62	Acceptable	SNHL	.	>=10
Overall	-1.20	-1.50	-0.89				

Model: Random-effects model
Heterogeneity: Tau-squared = 0.02, H-squared = 1.15, I-squared = 0.13
Homogeneity: Q = 4.65, df = 4, p-value = 0.33
Test of overall effect size: z = -7.79, p-value = 0.00

SUPPLEMENTARY DIGITAL MATERIAL 7

Supplementary Figure 5.—Differences in performance on *reach test* between *cases* and controls.

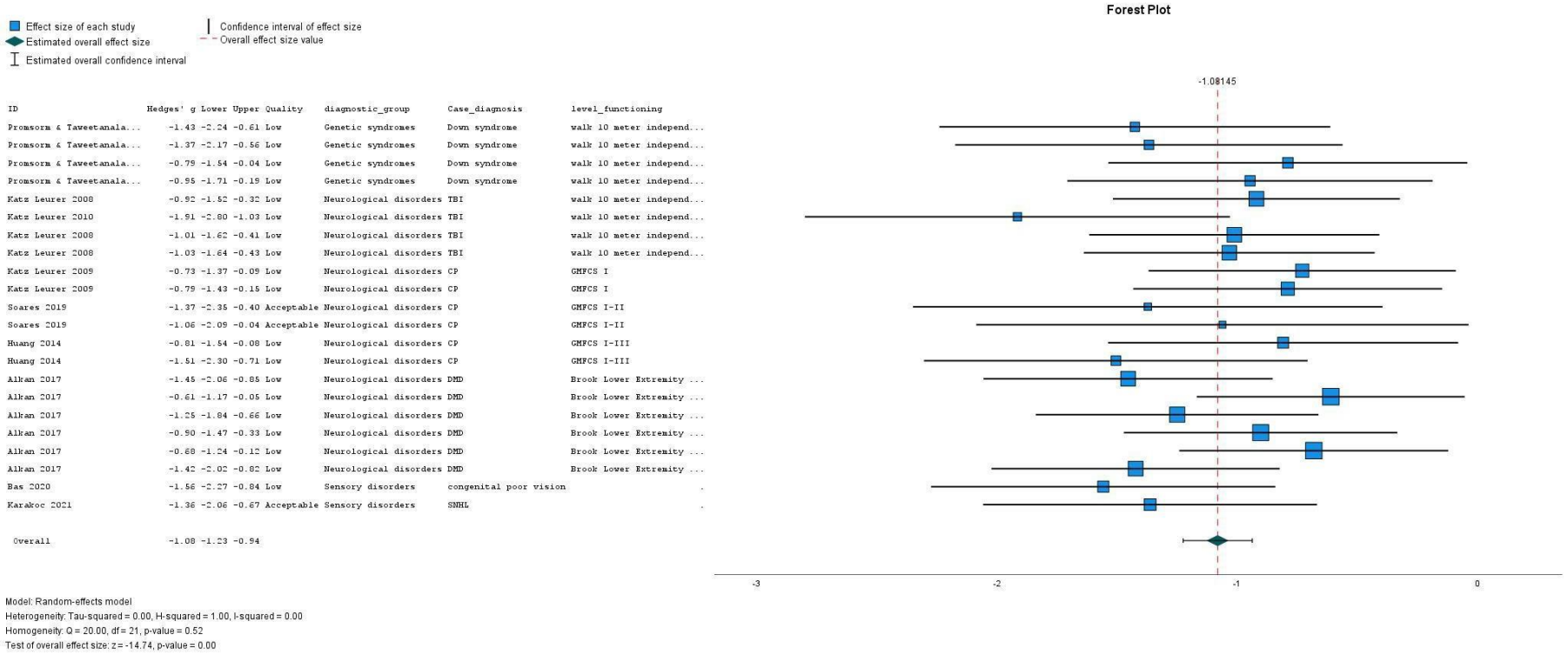


diagnostic_group	ID	Hedges' g	Lower	Upper	Quality	Case_diagnosis	level_functioning	agegroup
Genetic syndromes	Jain 2022	-0.65	-1.41	0.11	Low	Down syndrome	stand independently	>=10
	Jain 2022	-0.71	-1.48	0.06	Low	Down syndrome	stand independently	>=10
	Jain 2022	-0.91	-1.69	-0.12	Low	Down syndrome	stand independently	>=10
	Promsorn & Taweetanala...	-1.43	-2.24	-0.61	Low	Down syndrome	walk 10 meter independ...	7-9
	Promsorn & Taweetanala...	-1.37	-2.17	-0.56	Low	Down syndrome	walk 10 meter independ...	7-9
	Promsorn & Taweetanala...	-0.79	-1.54	-0.04	Low	Down syndrome	walk 10 meter independ...	7-9
	Promsorn & Taweetanala...	-0.95	-1.71	-0.19	Low	Down syndrome	walk 10 meter independ...	7-9
	Promsorn & Taweetanala...	-1.60	-2.44	-0.76	Low	Down syndrome	walk 10 meter independ...	>=10
	Promsorn & Taweetanala...	-0.51	-1.24	0.22	Low	Down syndrome	walk 10 meter independ...	>=10
	Promsorn & Taweetanala...	-0.70	-1.44	0.04	Low	Down syndrome	walk 10 meter independ...	>=10
	Promsorn & Taweetanala...	-0.69	-1.43	0.05	Low	Down syndrome	walk 10 meter independ...	>=10
	Subgroup Overall	-0.91	-1.14	-0.68				
Neurological disorders	Alkan 2017	-1.45	-2.06	-0.85	Low	DMD	Brook Lower Extremity ...	7-9
	Alkan 2017	-0.61	-1.17	-0.05	Low	DMD	Brook Lower Extremity ...	7-9
	Alkan 2017	-1.25	-1.84	-0.66	Low	DMD	Brook Lower Extremity ...	7-9
	Alkan 2017	-0.90	-1.47	-0.33	Low	DMD	Brook Lower Extremity ...	7-9
	Alkan 2017	-0.68	-1.24	-0.12	Low	DMD	Brook Lower Extremity ...	7-9
	Alkan 2017	-1.42	-2.02	-0.82	Low	DMD	Brook Lower Extremity ...	7-9
	Huang 2014	-0.81	-1.54	-0.08	Low	CP	GMFCS I-III	7-9
	Katz Leurer 2008	-0.92	-1.52	-0.32	Low	TBI	walk 10 meter independ...	7-9
	Katz Leurer 2009	-0.73	-1.37	-0.09	Low	CP	GMFCS I	7-9
	Katz Leurer 2009	-0.79	-1.43	-0.15	Low	CP	GMFCS I	7-9
	Katz Leurer 2010	-1.91	-2.80	-1.03	Low	TBI	walk 10 meter independ...	7-9
	Soares 2019	-1.37	-2.35	-0.40	Acceptable	CP	GMFCS I-II	7-9
	Soares 2019	-1.06	-2.09	-0.04	Acceptable	CP	GMFCS I-II	7-9
	Zaino 2004	-4.22	-5.50	-2.94	Low	CP	GMFCS I	>=10
	Zaino 2004	-6.74	-8.48	-4.99	Low	CP	GMFCS II	>=10
	Huang 2014	-1.51	-2.30	-0.71	Low	CP	GMFCS I-III	7-9
	Huang 2014	-1.81	-2.65	-0.97	Low	CP	GMFCS I-III	>=10
	Katz Leurer 2008	-1.01	-1.62	-0.41	Low	TBI	walk 10 meter independ...	7-9
	Katz Leurer 2008	-1.03	-1.64	-0.43	Low	TBI	walk 10 meter independ...	7-9
	Subgroup Overall	-1.45	-1.95	-0.96				
Respiratory disorders	Dik 2020	-1.28	-2.03	-0.53	Acceptable	cystic fibrosis	.	>=10
	Subgroup Overall	-1.28	-2.03	-0.53				
Sensory disorders	Bas 2020	-1.56	-2.27	-0.84	Low	congenital poor vision	.	7-9
	Karakoc 2021	-1.36	-2.06	-0.67	Acceptable	SNHL	.	7-9
	Subgroup Overall	-1.46	-1.96	-0.96				
Overall		-1.19	-1.41	-0.98				

Model: Random-effects model
Heterogeneity: Tau-squared = 0.24, H-squared = 2.83, I-squared = 0.65
Homogeneity: Q = 94.62, df = 32, p-value = 0.00
Test of overall effect size: z = -10.95, p-value = 0.00
Test of between-subgroup homogeneity: Q = 6.72, df = 3, p-value = 0.08

SUPPLEMENTARY DIGITAL MATERIAL 8

Supplementary Figure 6.—Subgroup analysis - Differences in performance on *reach test* between cases and controls in the age group 7-9 years old.

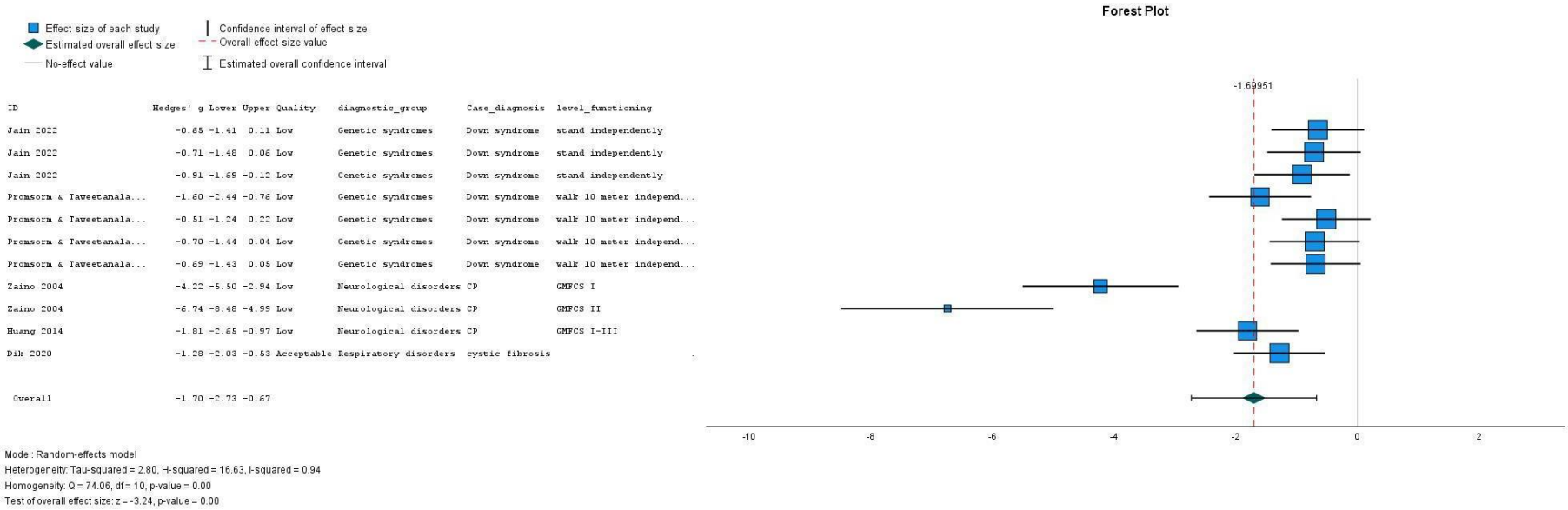


ID	Hedges' g	Lower	Upper	Quality	diagnostic_group	Case_diagnosis	level_functioning
Promsorn & Taweetanala...	-1.43	-2.24	-0.61	Low	Genetic syndromes	Down syndrome	walk 10 meter independ...
Promsorn & Taweetanala...	-1.37	-2.17	-0.56	Low	Genetic syndromes	Down syndrome	walk 10 meter independ...
Promsorn & Taweetanala...	-0.79	-1.54	-0.04	Low	Genetic syndromes	Down syndrome	walk 10 meter independ...
Promsorn & Taweetanala...	-0.95	-1.71	-0.19	Low	Genetic syndromes	Down syndrome	walk 10 meter independ...
Katz Leurer 2008	-0.92	-1.52	-0.32	Low	Neurological disorders	TBI	walk 10 meter independ...
Katz Leurer 2010	-1.91	-2.80	-1.03	Low	Neurological disorders	TBI	walk 10 meter independ...
Katz Leurer 2008	-1.01	-1.62	-0.41	Low	Neurological disorders	TBI	walk 10 meter independ...
Katz Leurer 2008	-1.03	-1.64	-0.43	Low	Neurological disorders	TBI	walk 10 meter independ...
Katz Leurer 2009	-0.73	-1.37	-0.09	Low	Neurological disorders	CP	GMFCS I
Katz Leurer 2009	-0.79	-1.43	-0.15	Low	Neurological disorders	CP	GMFCS I
Soares 2019	-1.37	-2.35	-0.40	Acceptable	Neurological disorders	CP	GMFCS I-II
Soares 2019	-1.06	-2.09	-0.04	Acceptable	Neurological disorders	CP	GMFCS I-II
Huang 2014	-0.81	-1.54	-0.08	Low	Neurological disorders	CP	GMFCS I-III
Huang 2014	-1.51	-2.30	-0.71	Low	Neurological disorders	CP	GMFCS I-III
Alkan 2017	-1.45	-2.06	-0.85	Low	Neurological disorders	DMD	Brook Lower Extremity ...
Alkan 2017	-0.61	-1.17	-0.05	Low	Neurological disorders	DMD	Brook Lower Extremity ...
Alkan 2017	-1.25	-1.84	-0.66	Low	Neurological disorders	DMD	Brook Lower Extremity ...
Alkan 2017	-0.90	-1.47	-0.33	Low	Neurological disorders	DMD	Brook Lower Extremity ...
Alkan 2017	-0.68	-1.24	-0.12	Low	Neurological disorders	DMD	Brook Lower Extremity ...
Alkan 2017	-1.42	-2.02	-0.82	Low	Neurological disorders	DMD	Brook Lower Extremity ...
Bas 2020	-1.56	-2.27	-0.84	Low	Sensory disorders	congenital poor vision	.
Karakoc 2021	-1.36	-2.06	-0.67	Acceptable	Sensory disorders	SNHL	.
Overall	-1.08	-1.23	-0.94				

Model: Random-effects model
Heterogeneity: Tau-squared = 0.00, H-squared = 1.00, I-squared = 0.00
Homogeneity: Q = 20.00, df = 21, p-value = 0.52
Test of overall effect size: z = -14.74, p-value = 0.00

SUPPLEMENTARY DIGITAL MATERIAL 9

Supplementary Figure 7.—Subgroup analysis - Differences in performance on *reach test* between cases and controls in the age group 10 years and older.

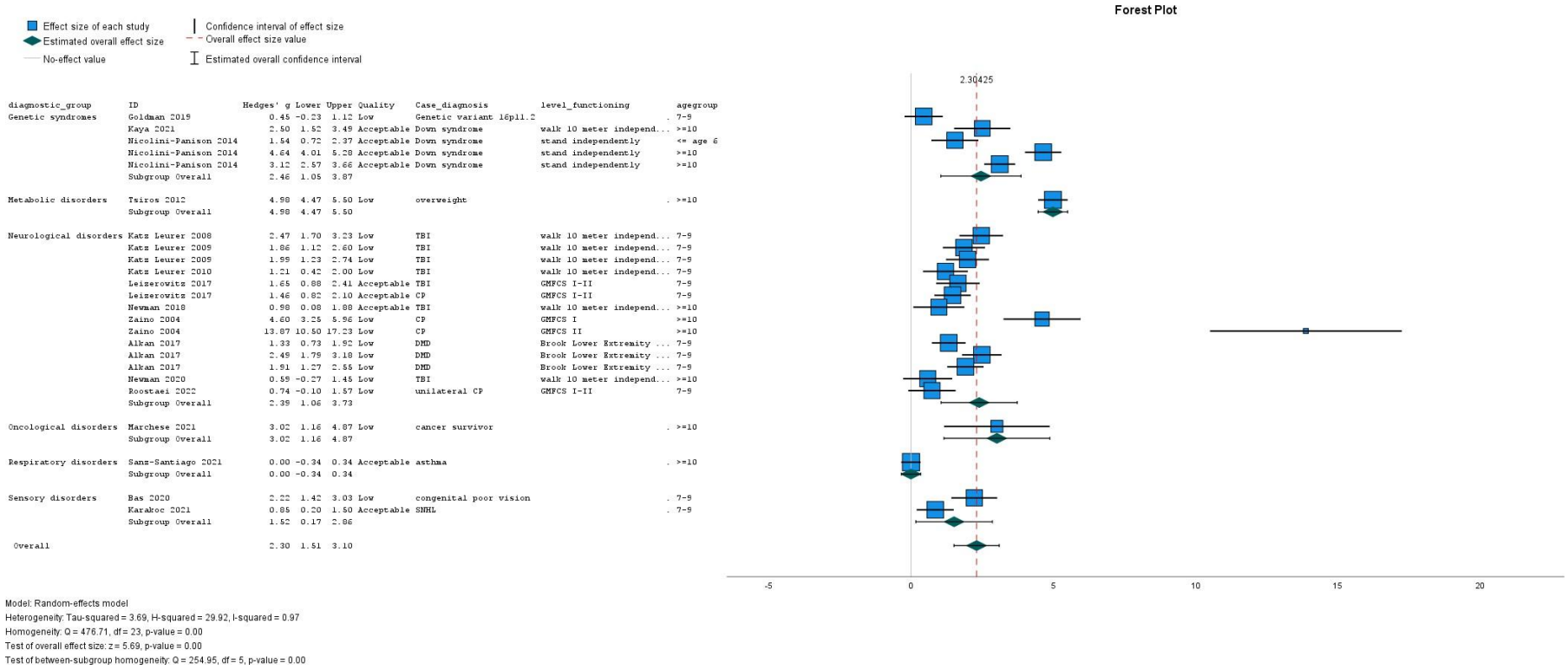


ID	Hedges' g	Lower	Upper	Quality	diagnostic_group	Case_diagnosis	level_functioning
Jain 2022	-0.65	-1.41	0.11	Low	Genetic syndromes	Down syndrome	stand independently
Jain 2022	-0.71	-1.48	0.06	Low	Genetic syndromes	Down syndrome	stand independently
Jain 2022	-0.91	-1.69	-0.12	Low	Genetic syndromes	Down syndrome	stand independently
Promsorn & Taweetanala...	-1.60	-2.44	-0.76	Low	Genetic syndromes	Down syndrome	walk 10 meter independ...
Promsorn & Taweetanala...	-0.51	-1.24	0.22	Low	Genetic syndromes	Down syndrome	walk 10 meter independ...
Promsorn & Taweetanala...	-0.70	-1.44	0.04	Low	Genetic syndromes	Down syndrome	walk 10 meter independ...
Promsorn & Taweetanala...	-0.69	-1.43	0.05	Low	Genetic syndromes	Down syndrome	walk 10 meter independ...
Zaino 2004	-4.22	-5.50	-2.94	Low	Neurological disorders	CP	GMFCS I
Zaino 2004	-6.74	-8.48	-4.99	Low	Neurological disorders	CP	GMFCS II
Huang 2014	-1.81	-2.65	-0.97	Low	Neurological disorders	CP	GMFCS I-III
Dik 2020	-1.28	-2.03	-0.53	Acceptable	Respiratory disorders	cystic fibrosis	
Overall	-1.70	-2.73	-0.67				

Model: Random-effects model
Heterogeneity: Tau-squared = 2.80, H-squared = 16.63, I-squared = 0.94
Homogeneity: Q = 74.06, df = 10, p-value = 0.00
Test of overall effect size: z = -3.24, p-value = 0.00

SUPPLEMENTARY DIGITAL MATERIAL 10

Supplementary Figure 8.—Differences in performance on the Timed Up and Go test between cases and controls.

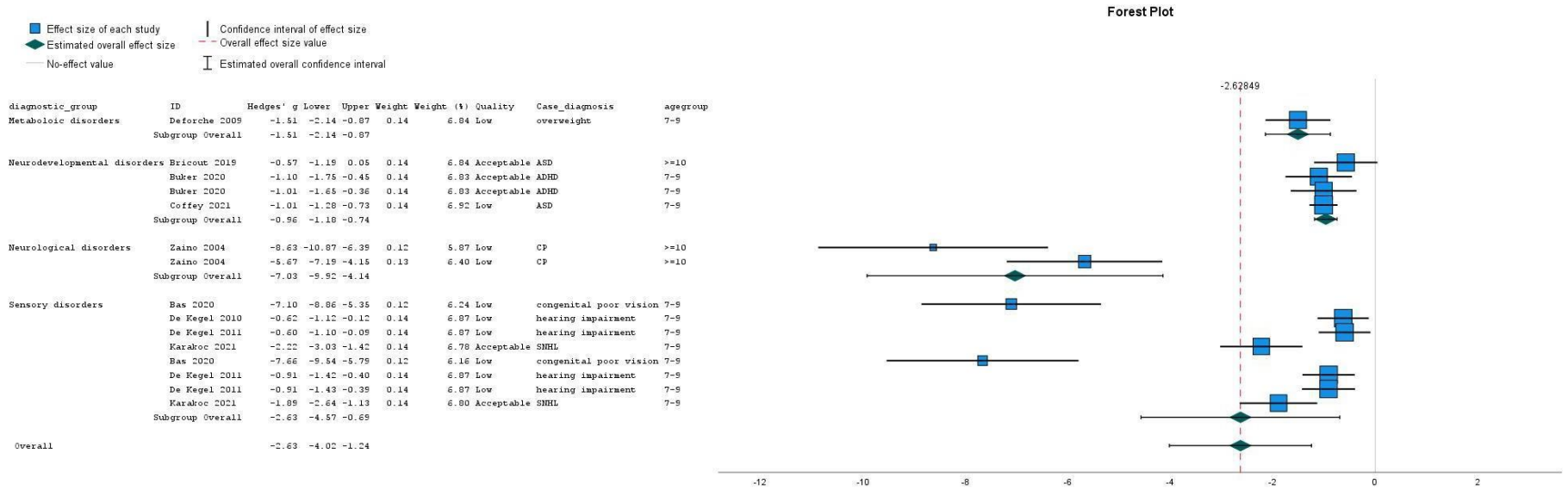


diagnostic_group	ID	Hedges' g	Lower	Upper	Quality	Case_diagnosis	level_functioning	agegroup
Genetic syndromes	Goldman 2019	0.45	-0.23	1.12	Low	Genetic variant 16p11.2	.	7-9
	Kaya 2021	2.50	1.52	3.49	Acceptable	Down syndrome	walk 10 meter independ...	>=10
	Nicolini-Panison 2014	1.54	0.72	2.37	Acceptable	Down syndrome	stand independently	<= age 6
	Nicolini-Panison 2014	4.64	4.01	5.28	Acceptable	Down syndrome	stand independently	>=10
	Nicolini-Panison 2014	3.12	2.57	3.66	Acceptable	Down syndrome	stand independently	>=10
	Subgroup Overall	2.46	1.05	3.87				
Metabolic disorders	Tsiros 2012	4.98	4.47	5.50	Low	overweight	.	>=10
	Subgroup Overall	4.98	4.47	5.50				
Neurological disorders	Katz Leurer 2008	2.47	1.70	3.23	Low	TBI	walk 10 meter independ...	7-9
	Katz Leurer 2009	1.86	1.12	2.60	Low	TBI	walk 10 meter independ...	7-9
	Katz Leurer 2009	1.99	1.23	2.74	Low	TBI	walk 10 meter independ...	7-9
	Katz Leurer 2010	1.21	0.42	2.00	Low	TBI	walk 10 meter independ...	7-9
	Leizerowitz 2017	1.65	0.88	2.41	Acceptable	TBI	GMFCS I-II	7-9
	Leizerowitz 2017	1.46	0.82	2.10	Acceptable	CP	GMFCS I-II	7-9
	Newman 2018	0.98	0.08	1.88	Acceptable	TBI	walk 10 meter independ...	>=10
	Zaino 2004	4.60	3.25	5.96	Low	CP	GMFCS I	>=10
	Zaino 2004	13.87	10.50	17.23	Low	CP	GMFCS II	>=10
	Alkan 2017	1.33	0.73	1.92	Low	DMD	Brook Lower Extremity ...	7-9
	Alkan 2017	2.49	1.79	3.18	Low	DMD	Brook Lower Extremity ...	7-9
	Alkan 2017	1.91	1.27	2.55	Low	DMD	Brook Lower Extremity ...	7-9
	Newman 2020	0.59	-0.27	1.45	Low	TBI	walk 10 meter independ...	>=10
	Roostaei 2022	0.74	-0.10	1.57	Low	unilateral CP	GMFCS I-II	7-9
	Subgroup Overall	2.39	1.06	3.73				
Oncological disorders	Marchese 2021	3.02	1.16	4.87	Low	cancer survivor	.	>=10
	Subgroup Overall	3.02	1.16	4.87				
Respiratory disorders	Sanz-Santiago 2021	0.00	-0.34	0.34	Acceptable	asthma	.	>=10
	Subgroup Overall	0.00	-0.34	0.34				
Sensory disorders	Bas 2020	2.22	1.42	3.03	Low	congenital poor vision	.	7-9
	Karakoc 2021	0.85	0.20	1.50	Acceptable	SNHL	.	7-9
	Subgroup Overall	1.52	0.17	2.86				
Overall		2.30	1.51	3.10				

Model: Random-effects model
Heterogeneity: Tau-squared = 3.69, H-squared = 29.92, I-squared = 0.97
Homogeneity: Q = 476.71, df = 23, p-value = 0.00
Test of overall effect size: z = 5.69, p-value = 0.00
Test of between-subgroup homogeneity: Q = 254.95, df = 5, p-value = 0.00

SUPPLEMENTARY DIGITAL MATERIAL 11

Supplementary Figure 9.—Differences in performance on *One Leg Stance* between *cases* and controls.

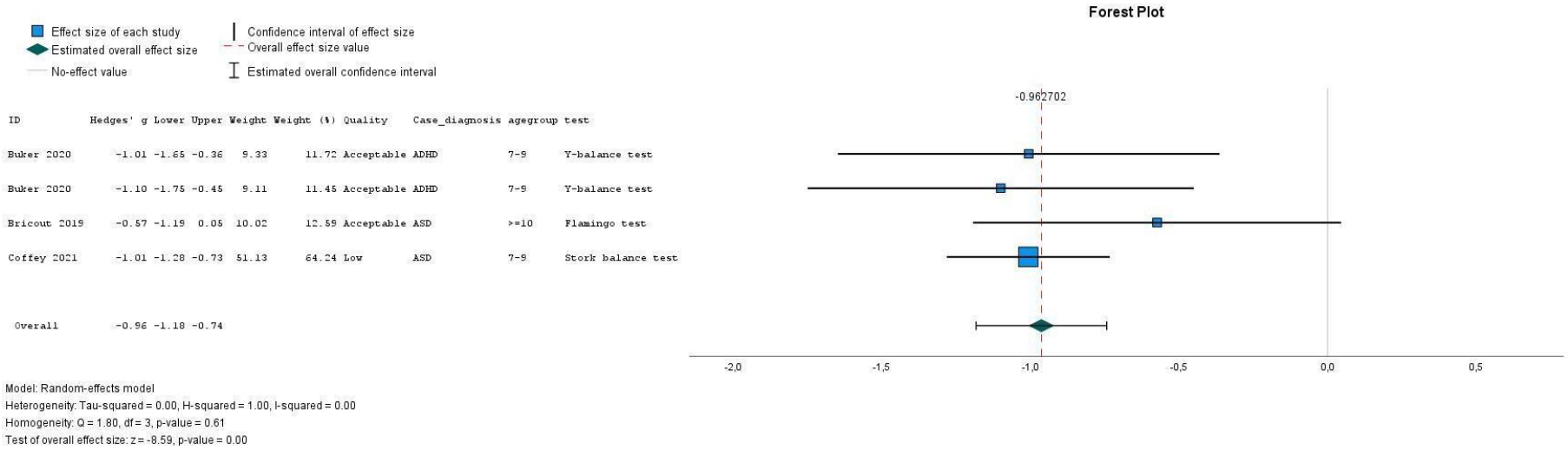


diagnostic_group	ID	Hedges' g	Lower	Upper	Weight	Weight (%)	Quality	Case_diagnosis	agegroup
Metaboloic disorders	Deforche 2009	-1.51	-2.14	-0.87	0.14	6.84	Low	overweight	7-9
	Subgroup Overall	-1.51	-2.14	-0.87					
Neurodevelopmental disorders	Bricout 2019	-0.57	-1.19	0.05	0.14	6.84	Acceptable	ASD	>=10
	Buker 2020	-1.10	-1.75	-0.45	0.14	6.83	Acceptable	ADHD	7-9
	Buker 2020	-1.01	-1.65	-0.36	0.14	6.83	Acceptable	ADHD	7-9
	Coffey 2021	-1.01	-1.28	-0.73	0.14	6.92	Low	ASD	7-9
	Subgroup Overall	-0.96	-1.18	-0.74					
Neurological disorders	Zaino 2004	-8.63	-10.87	-6.39	0.12	5.87	Low	CP	>=10
	Zaino 2004	-5.67	-7.19	-4.15	0.13	6.40	Low	CP	>=10
	Subgroup Overall	-7.03	-9.92	-4.14					
Sensory disorders	Bas 2020	-7.10	-8.86	-5.35	0.12	6.24	Low	congenital poor vision	7-9
	De Kegel 2010	-0.62	-1.12	-0.12	0.14	6.87	Low	hearing impairment	7-9
	De Kegel 2011	-0.60	-1.10	-0.09	0.14	6.87	Low	hearing impairment	7-9
	Karakoc 2021	-2.22	-3.03	-1.42	0.14	6.78	Acceptable	SNHL	7-9
	Bas 2020	-7.66	-9.54	-5.79	0.12	6.16	Low	congenital poor vision	7-9
	De Kegel 2011	-0.91	-1.42	-0.40	0.14	6.87	Low	hearing impairment	7-9
	De Kegel 2011	-0.91	-1.43	-0.39	0.14	6.87	Low	hearing impairment	7-9
	Karakoc 2021	-1.89	-2.64	-1.13	0.14	6.80	Acceptable	SNHL	7-9
	Subgroup Overall	-2.63	-4.57	-0.69					
Overall		-2.63	-4.02	-1.24					

Model: Random-effects model
Heterogeneity: Tau-squared = 7.22, H-squared = 74.32, I-squared = 0.99
Homogeneity: Q = 194.22, df = 14, p-value = 0.00
Test of overall effect size: z = -3.71, p-value = 0.00
Test of between-subgroup homogeneity: Q = 21.59, df = 3, p-value = 0.00

SUPPLEMENTARY DIGITAL MATERIAL 12

Supplementary Figure 10.—Subgroup analysis - Differences in performance on One Leg Stance between children with neurodevelopmental disorders and controls.

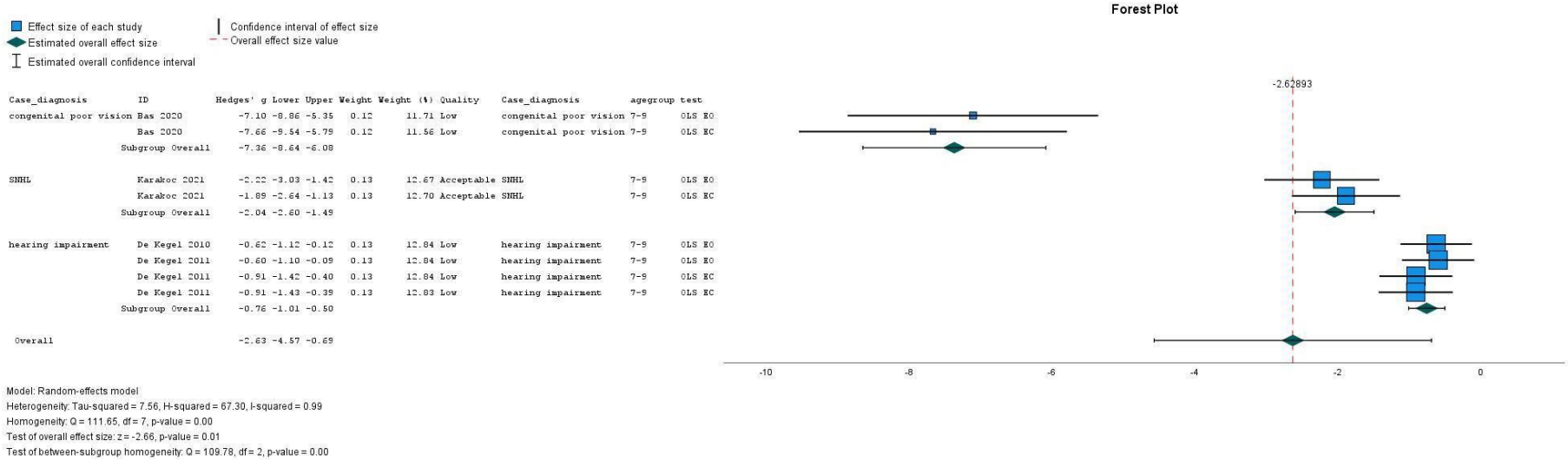


ID	Hedges' g	Lower	Upper	Weight	Weight (%)	Quality	Case_diagnosis	agegroup	test
Buker 2020	-1.01	-1.65	-0.36	9.33	11.72	Acceptable	ADHD	7-9	Y-balance test
Buker 2020	-1.10	-1.75	-0.45	9.11	11.45	Acceptable	ADHD	7-9	Y-balance test
Bricout 2019	-0.57	-1.19	0.05	10.02	12.59	Acceptable	ASD	>=10	Flamingo test
Coffey 2021	-1.01	-1.28	-0.73	51.13	64.24	Low	ASD	7-9	Stork balance test
Overall	-0.96	-1.18	-0.74						

Model: Random-effects model
Heterogeneity: Tau-squared = 0.00, H-squared = 1.00, I-squared = 0.00
Homogeneity: Q = 1.80, df = 3, p-value = 0.61
Test of overall effect size: z = -8.59, p-value = 0.00

SUPPLEMENTARY DIGITAL MATERIAL 13

Supplementary Figure 11.— Subgroup analysis - Differences in performance on One Leg Stance between children with sensory deficits and controls.



Case_diagnosis	ID	Hedges' g	Lower	Upper	Weight	Weight (%)	Quality	Case_diagnosis	agegroup	test
congenital poor vision	Bas 2020	-7.10	-8.86	-5.35	0.12	11.71	Low	congenital poor vision	7-9	OLS EO
	Bas 2020	-7.66	-9.54	-5.79	0.12	11.56	Low	congenital poor vision	7-9	OLS EC
	Subgroup Overall	-7.36	-8.64	-6.08						
SNHL	Karakoc 2021	-2.22	-3.03	-1.42	0.13	12.67	Acceptable	SNHL	7-9	OLS EO
	Karakoc 2021	-1.89	-2.64	-1.13	0.13	12.70	Acceptable	SNHL	7-9	OLS EC
	Subgroup Overall	-2.04	-2.60	-1.49						
hearing impairment	De Kegel 2010	-0.62	-1.12	-0.12	0.13	12.84	Low	hearing impairment	7-9	OLS EO
	De Kegel 2011	-0.60	-1.10	-0.09	0.13	12.84	Low	hearing impairment	7-9	OLS EO
	De Kegel 2011	-0.91	-1.42	-0.40	0.13	12.84	Low	hearing impairment	7-9	OLS EC
	De Kegel 2011	-0.91	-1.43	-0.39	0.13	12.83	Low	hearing impairment	7-9	OLS EC
	Subgroup Overall	-0.76	-1.01	-0.50						
Overall		-2.63	-4.57	-0.69						

Model: Random-effects model
Heterogeneity: Tau-squared = 7.56, H-squared = 67.30, I-squared = 0.99
Homogeneity: Q = 111.65, df = 7, p-value = 0.00
Test of overall effect size: z = -2.66, p-value = 0.01
Test of between-subgroup homogeneity: Q = 109.78, df = 2, p-value = 0.00