

Tools and Techniques Used With Robotic Devices to Quantify Upper-Limb Function in Typically Developing Children: A Systematic Review

Stephan CD Dobri¹, Hana M Ready and Theresa Claire Davies¹

Department of Mechanical and Materials Engineering, Queen's University, Canada.

Rehabilitation Process and Outcome
Volume 9: 1–17
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1179572720979013



ABSTRACT

BACKGROUND: Robotic devices have been used to quantify function, identify impairment, and rehabilitate motor function extensively in adults, but less-so in younger populations. The ability to perform motor actions improves as children grow. It is important to quantify this rate of change of the neurotypical population before attempting to identify impairment and target rehabilitation techniques.

OBJECTIVES: For a population of typically developing children, this systematic review identifies and analyzes tools and techniques used with robotic devices to quantify upper-limb motor function. Since most of the papers also used robotic devices to compare function of neurotypical to pathological populations, a secondary objective was introduced to relate clinical outcome measures to identified robotic tools and techniques.

METHODS: Five databases were searched between February 2019 and August 2020, and 226 articles were found, 19 of which are included in the review.

RESULTS: Robotic devices, tasks, outcome measures, and clinical assessments were not consistent among studies from different settings but were consistent within laboratory groups. Fifteen of the 19 articles evaluated both typically developing and pathological populations.

CONCLUSION: To optimize universally comparable outcomes in future work, it is recommended that a standard set of tasks and measures is used to assess upper-limb motor function. Standardized tasks and measures will facilitate effective rehabilitation.

KEYWORDS: Robotics, upper-limb, children, quantify, quantification, motor function, review, typically developing

RECEIVED: May 26, 2020. **ACCEPTED:** November 16, 2020.

TYPE: Systematic Review

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported through funding from the Natural Sciences and Engineering Research Council grant number 513272-17, NSERC RGPIN 2016-04669, and the Ontario Graduate Scholarship (OGS).

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Stephan CD Dobri, Department of Mechanical and Materials Engineering, Queen's University, McLaughlin Hall 130 Stuart Street, Kingston, Ontario K7L 2V9, Canada. Email: stephan.dobri@queensu.ca

Introduction

Robotic devices can quantify characteristics of upper-limb function, identify indicators of impairment, and assist in rehabilitation.¹⁻³ Repeatable tasks with a high degree of accuracy can be conducted to test various aspects of a person's ability, from motor function to short term memory.^{2,4} In adult populations, robotic devices have been used extensively to quantify typical age-related changes, as well as deficits caused by injuries or illnesses.^{2,4-7} Additionally, robots have been used for rehabilitation in adults, often for those who have experienced stroke.²

Robotic devices offer the same testing capabilities for younger populations but have not been used to the same extent. For example, 2 reviews of robotic interaction, 1 examining the adult population with stroke, and the other, the child population with cerebral palsy, resulted in 28 versus 9 articles, respectively.^{2,3}

Robotic quantification of the adult population is straightforward as motor control parameters across the ages from 30 to 60 years display little variation.⁸ To identify pathological differences in motor control, the clinician or researcher can readily compare robotic outcome measures to a general database of

members within the typical population. However, as children grow, motor control improves, making it difficult to generalize the parameters associated with each age. It is important to understand how function changes across the developmental stages from 5 to 18 years of age, to better identify anomalies related to pathology. Additionally, the goal of therapy is to return function to levels of typical development, and the age at which each developmental goal would typically be achieved must be identified.

The primary objective of this review was to identify the robotic tools and measures that have been used to quantify upper-limb function in typically developing children and youth. Since most papers also used robotic devices to compare function of neurotypical to pathological populations, a secondary objective was introduced to relate clinical outcome measures to robotic tools and techniques to allow comparison with pathological populations. Knowing which tools and measures are most clinically relevant can enable clinicians and researchers to choose the best methods of evaluating motor control of children with neurological conditions as compared to those who are typically developing.



Methods

Study design

This study is a systematic review that followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.

Eligibility criteria

To be included, articles had to: be written in English, published in peer-reviewed journals, and discuss the use of a robotic apparatus to assess upper-limb function in typically developing children. Articles were excluded if: they assessed only a pathological population, the mean age of participants was above 18 years of age, only lower-limb function was assessed, or the robotic device/system was evaluated rather than testing function of children. To ensure a comprehensive review, there were no restrictions on publication date, study design, or types of outcomes/asures of upper-limb function. References in any review articles found within the search were used to verify that all relevant articles were included.

Information sources and search

To identify research that used a robotic apparatus to quantify upper-limb motor function in children and adolescents, a search was conducted between February 2019 and August 2020 among 5 databases (EMBASE, Engineering Village, OVID MEDLINE, PubMed, and Web of Science). The search was performed with the following search string: (Upper-Limb OR Upper—Limb OR Arm OR Upper-Extremity* OR Upper—Extremity*) AND ((Robot* OR Exoskeleton) AND (Test* OR Assessment OR Quantification OR Evaluation OR Examination)) AND ((Healthy OR Typical* Develop* OR Control OR Normal) AND (Child* OR Youth OR Adolescent OR School-Age OR Pediatric OR Teen)).

Study selection

The titles were first searched to eliminate any articles that did not meet the inclusion criteria, followed by abstract review. The titles and abstracts were reviewed by 1 reviewer. Typically, at least 2 reviewers would be involved at each stage to reduce the risk of bias in paper selection. To help reduce the chances of bias from 1 reviewer screening titles and abstracts, the reviewer was instructed to be lenient when judging the exclusion criteria. Leniency at this stage resulted in a greater number of full text articles being reviewed by multiple reviewers than if multiple reviewers had been involved at this stage in screening.

For each paper that was included after the abstract review,² reviewers assessed the papers using a modified version of the McMaster Quantitative Review Form (<https://srs-mcmaster.ca/research/evidence-based-practice-research-group/>). This form was used to identify any potential biases (ie, subject selection, measurement, and intervention biases), study design,

study participants, ethics procedure, outcome measures, intervention, and study results. The form was also used to extract and record data about the measures and outcomes that were reported in each study. The agreement between the reviewers for article inclusion after full paper assessment was calculated using the Cohen's kappa statistic. When reviewers disagreed whether to include an article, the reasons were discussed with the third reviewer and a decision on inclusion was reached.

Results

A total of 476 articles were identified from the initial search of databases. After removing duplicates, 226 articles remained and were assessed. After the complete selection process, 19 were included in this review. Cohen's kappa was found to be 0.76, indicating substantial agreement between the investigators. The selection process is illustrated in Figure 1.

Data about the robotic apparatus, tests of function, and outcome measures for the tests were recorded and summarized. Additionally, the demographic information of participant groups was recorded to compare among studies.

After reviewing all articles, general themes were established and summarized. It is important to note that due to differences in methods and populations studied, a meta-analysis could not be completed as there was insufficient replication of parameters. Correlations between clinical and robotic measures of function could have been combined in a meta-analysis had there been sufficient replication among studies.

Study populations and study design

The demographics and the study design of each study are summarized in brief in Table 1.

Most of the articles included within this review discussed the study of typically developing children as compared to pathological populations (15 of 19 articles). In 11 of these articles, children with cerebral palsy provided the comparator.^{11,12,15-22,27} Four articles each evaluated a different pathological population; children with Friedreich's Ataxia, Fetal Alcohol Spectrum Disorder (FASD), Autism Spectrum Disorder (ASD), and Developmental Coordination Disorder (DCD).^{9,13,24,26} The final 4 studied only typically developing populations.^{10,14,23,25}

Of the included articles, 6 had sample sizes of 10 participants or fewer in at least 1 of the populations studied,^{9,10,12,22-24} 8 had sample sizes of 50 participants or more in at least 1 of the populations studied,^{14-18,20,24,26} with 2 articles having over 200 participants in the control group.^{12,24} For larger sample sizes, the authors could make reliable group comparisons, whereas smaller sample sizes required authors to acknowledge the limitations of their results. Only 1 article justified the sample size with a priori power analysis, while one other completed a power analysis of their results.^{17,25}

Only 2 articles specifically stated age- and/or sex-matched control populations were used to compare against pathological populations.^{13,22} All other articles ensured that control

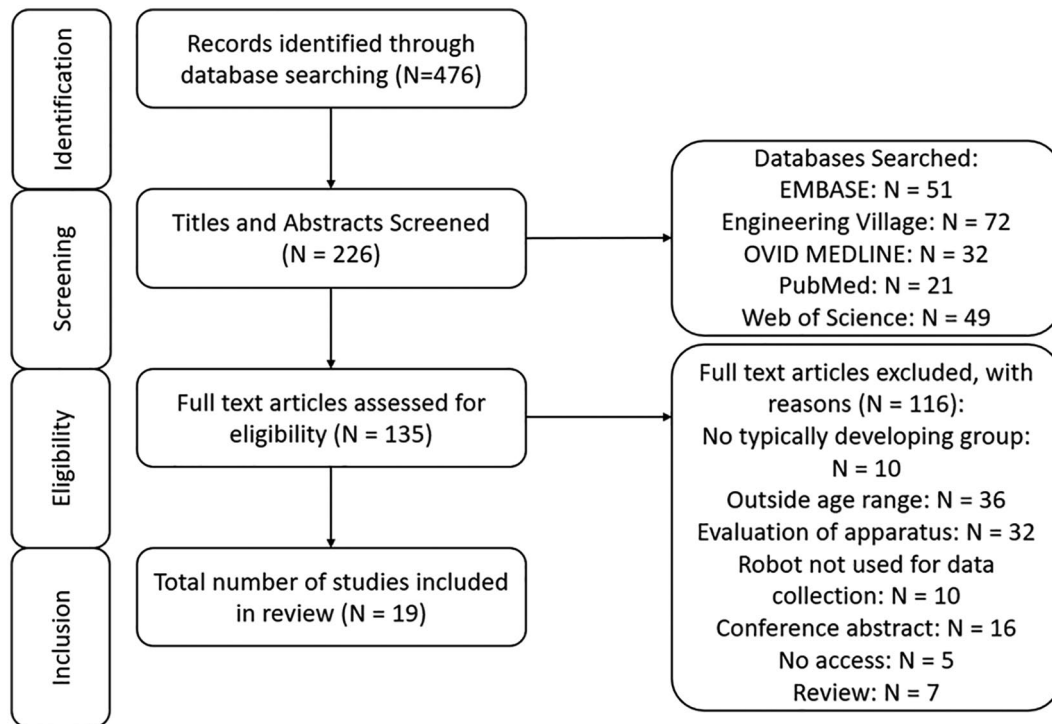


Figure 1. Flowchart showing article selection process.

populations and pathological populations had the same age range. The 2 articles with age- and/or sex-matched controls also had smaller population sizes (7 and 18 controls, respectively).^{13,22}

Three articles explicitly stated that data was separated by participant sex or indicated that the analysis accounted for sex.^{12,24,26} Only 4 articles^{12,22,24,26} considered the effect of participant sex on performance, including the article with age- and sex-matched controls.^{12,22,24,26} One of the articles explicitly tested for significant differences in male and female performance and did not find significant differences in any of the parameters due to sex.²⁶

One article was a randomized control trial,¹⁰ with all others being cross-sectional studies.

Robotic devices used

There were 4 robotic devices used for evaluation of motor control, each of which is briefly summarized in Table 2. All articles discussed a robot that was either already commercially available or, in the case of the MIT Manus, the version of the robot before it was commercialized. One of the 4 robots can enable evaluation with a 3D workspace, with 6 degrees of freedom (DOF) (the PHANToM), while the others work in a planar, 2D workspace with only 2 DOF. These differences in workspaces and DOF restrict each robot to specific tasks and measurements of function. Additionally, only 1 of the 4 robots (the Kinarm Exoskeleton Robot Lab) can test both upper-limbs simultaneously (bimanual tests), while the others only focus on unimanual tasks.

Aspects of function being assessed with robotic tests and measures

All aspects of function, robotic tests, robotic measures, and clinical measures used in each article are summarized in Table 3. Thirteen articles compared robotic measures of performance to clinical outcome measures.^{10,11,13-22,27} The definitions and descriptions of each robotic and clinical outcome measure can be found in their respective articles.

Overall motor performance was most commonly assessed, and occurred in 11 articles.^{10-16,20,21,23,26} Evaluation of position sense was reported in 4 articles,^{10,17,19,27} while both motor adaptation/learning^{9,22,25} and kinesthesia^{10,18,19} were each assessed in 3 articles.

Robotic tests and measures of function

Eight articles included reports of both arms,^{10,12,15,16,20,21,24,26} 9 assessed only 1 arm,^{11,13,14,17-19,23,25,27} and 2 did not specify.^{9,22} Two articles found a significant difference in performance between dominant and non-dominant arms,^{20,26} while the other 6 articles that tested both arms did not state significance.^{10,12,15,16,21,24}

Point-to-point reaching tasks (without perturbations) were the most common task type, appearing in 10 articles.^{10-14,16,20,21,23,26} The point-to-point reaching task was the only task that had a similar experimental paradigm among different laboratory groups and robots. Generally, a virtual target would appear, and the participant would reach toward the target. The target distances and orientation, as well as the calculations of parameters, differed among robotic platforms.

Table 1. Study design (StD) and summary of all participant group demographics for included articles. Study design is identified as CS for cross-sectional study or RCT for randomized control study.

ARTICLE (STD)	TYPICALLY DEVELOPING GROUP	PATHOLOGICAL OR INTERVENTION POPULATION
Casellato et al, ⁹ (CS)	16 total (7F, 9M). 8 to 10years old. 3L handed; 13R handed.	ASD: 1 M, 8years old, R handed.
Cole et al, ¹⁰ (RCT)	8 total (3F, 5M). 12 to 18years old (mean age: 15.8, SD: 1.3). All R handed. Both arms tested.	Pre- and post- intervention with transcranial direct current stimulation (tDCS) or high definition (HD)-tDCS (typically developing populations with an intervention): tDCS: 8 total (6F, 2M). 12 to 18years old (mean age: 15.9, SD: 1.5). All R handed. HD-tDCS: 8 total (4F, 4M). 12 to 18years old (mean age: 14.8, SD: 1.7). All R handed. Both arms tested.
Dehem et al, ¹¹ (CS)	49 total (28F, 21 M). 6 to 12years old (mean age: 9.0, SD: 2.02). 7L handed; 42R handed. Dominant arm tested.	Unilateral hemiparetic CP: 20 total (11F, 9M). 6 to 12years old (mean age: 10.0, SD: 1.93). MACS level I-III. 5L affected; 15R affected. Affected arm tested.
Dobri et al, ¹² (CS)	288 total (98F, 190M). 5 to 18years old (mean age: 12.9, SD: 3.2). Both arms tested.	CP: 3 total (1F, 2M). 10, 11, and 12years old. Both arms tested. All MACS level and GMFCS level 1.
Germanotta et al, ¹³ (CS)	18 total (13 F, 5 M). 7 to 28years old (mean age: 15.1). Age and sex matched to pathological population. Dominant arm tested.	Friedreich's Ataxia: 14 total (10F, 4M). 6 to 28years old (mean age: 15.3). 1 L handed; 13R handed. Dominant arm tested.
Gilliaux et al, ¹⁴ (CS)	93 total (52F, 41 M). 3 to 12years old (mean age: 7.8, SD: 2.7). 10L handed; 83 R handed. Dominant arm tested.	None
Hawe et al, ¹⁵ (CS)	146 total (69F, 77M). 6 to 19years old (mean age: 13.0, SD: 4.1). 12L handed, 134R handed. Both arms tested simultaneously.	Hemiparetic CP: AIS and PVI AIS: 26 total (9F, 17M). 6 to 19years old (mean age: 12.8, SD: 4.0). MACS level I-II. 8 L affected, 18 R affected. PVI: 19 total (6F, 13M). 6 to 19years old (mean age: 11.6, SD: 3.7). MACS level I-II. 10L affected, 9 R affected. Both arms tested simultaneously.
Hawe et al, ¹⁶ (CS)	155 total (74F, 81M). 6 to 19years old (mean age: 12.5, SD: 4.0). 13L handed, 141 R handed, 1 ambidextrous. Both arms tested simultaneously.	Hemiparetic CP: AIS and PVI AIS: 28 total (10F, 18M). 6 to 19years old (mean age: 12.4, SD: 4.0). MACS level I-II. 9L affected, 19R affected. PVI: 21 total (6F, 15M). 6 to 19years old (mean age: 11.7, SD: 3.8). MACS level I-II. 11 L affected, 10R affected. Both arms tested simultaneously.
Kuczynski et al, ¹⁷ (CS)	60 total (22F, 38M). 6 to 19years old (mean age: 12.3, SD: 3.6). 4L handed; 56R handed.	Hemiparetic CP: AIS and PVI. AIS: 22 total (10F, 12M). 6 to 19years old (mean age: 13, SD: 3.6). 6 MACS level I, 12MACS level II; 4 unknown. 8L affected; 14 R affected. PVI: 18 total (7F, 11M). 6 to 19years old (mean age: 12.1, SD: 3.5). 8 MACS level I, 5 MACS level II; 5 unknown. 9L affected; 9 R affected. Unaffected arm tested.
Kuczynski et al, ¹⁸ (CS)	106 total (51F, 55M). 6 to 19years old (mean age: 12.3, SD: 3.9). 11L handed; 95R handed. Non-dominant arm tested.	Hemiparetic CP: AIS and PVI AIS: 23 total (10F, 13M). 6 to 19years old (mean age: 12.7, SD: 3.7). 4 MACS level I; 16 MACS level II; 3 unknown. 9L affected; 14 R affected. PVI: 20 total (8F, 12M). 6 to 19years old (mean age: 11.6, SD: 3.7). 8 MACS level I; 5 MACS level II; 7 unknown. 10 each L and R affected.

(Continued)

Table 1. (Continued)

ARTICLE (STD)	TYPICALLY DEVELOPING GROUP	PATHOLOGICAL OR INTERVENTION POPULATION
		Affected arm tested.
Kuczynski et al, ¹⁹ (CS)	21 total (9F, 12M). 6 to 19years old (mean age: 12.1, SD: 3.2). All R handed. Dominant arm tested.	Hemiparetic CP: AIS and PVI.
		AIS: 14 total (5F, 9M). 6 to 19years old (mean age: 12, SD: 3.7). MACS level I-IV. 9L affected; 5 R affected.
		PVI: 15 total (6F, 9M). 6 to 19years old (mean age: 12.1, SD: 3.3). 8L affected; 7 R affected. Unaffected arm tested.
Kuczynski et al, ²⁰ (CS)	147 total (71 F, 76 M). 6 to 19years old (mean age: 12.7, SD: 3.9). 8 L handed; 127 R handed; 12 mixed. Both arms tested.	Hemiparetic CP: AIS and PVI
		AIS: 28 total (10F, 18M). 6 to 19years old (mean age: 12.5, SD: 3.9). MACS level I-IV. 10L affected; 18 R affected.
		PVI: 22 total (8F, 14M). 6 to 19years old (mean age: 11.5, SD: 3.8). MACS level I-IV. 11 each L and R affected.
		Both arms tested.
Kuczynski et al, ²¹ (CS)	26 total (11 F, 15M). 6 to 19years old (mean age: 12.3, SD: 3.5). All R handed. Both arms tested.	Hemiparetic CP: AIS and PVI
		AIS: 17 total (5F, 12M). 6-19years old (mean age: 12.1, SD: 4.3). MACS level I-IV. 12L affected, 5 R affected.
		PVI: 16 total (6F, 10M). 6 to 19years old (mean age: 11.7, SD: 3.6). MACS level I-IV. 9L affected; 7 R affected.
		Both arms tested.
Masia et al, ²² (CS)	7 total. 8 to 14 years old (mean age: 9). Age-matched to pathological population.	Hemiparetic CP: 7 total (all M). 7 to 14 years old (mean age: 10.14). All R affected.
Pacilli et al, ²³ (CS)	10 children (7-10years old)	None
	10 adults (23-25 years old)	
	All R handed. Dominant arm tested.	
Skarsgard et al, ²⁴ (CS)	207 total (64F, 143M). 5 to 18years old (mean age: 13, SD: 3). Both arms tested.	DCD: 3 total (all M). 7, 8, and 10 years old. Both arms tested.
Takahashi et al, ²⁵ (CS)	Grouped ages.	None
	6 to 8years old: 17 total (7F, 10M)	
	9 to 12years old: 13 total (8F, 5M)	
	13 to 17years old: 13 total (6F, 7M)	
	>17years old: 12 total (6F, 6M)	
	Dominant arm tested.	
Williams et al, ²⁶ (CS)	152 total (79 F, 73 M). 5 to 18years old (mean age: 11.16, SD: 4.1). 16 L handed; 136 R handed. Both arms tested.	FASD: 31 total (12F, 19M). 5 to 18years old (mean age: 11.5, SD: 3.3). 10L handed; 21 R handed. Both arms tested.
Woodward et al, ²⁷ (CS)	21 total (10F, 11 M). 6 to 19 years old (mean age: 12.6, SD: 3.5). All R handed. Dominant arm tested.	Hemiparetic CP: PVI
		17 total (6F, 11M). 6 to 19years old (mean age: 12.2, SD: 4.0). 9L affected, 8R affected. Unaffected arm tested.

Handedness is identified as L for left, R for right.

Abbreviations: The following pathologies were identified: CP, cerebral palsy; AIS, Arterial Ischemic Stroke; PVI, Periventricular Venous Infarction; ASD, Autism Spectrum Disorder; FASD, Fetal Alcohol Spectrum Disorder; DCD, Developmental Coordination Disorder; Other acronyms: MACS, Manual Ability Classification System; GMFCS, Gross Motor Function Classification System.

Age is reported in years with SD as standard deviation.

Table 2. Information on the robots from articles included in review.

ROBOT	COMPANY	ROBOT WORKSPACE	NUMBER OF ARTICLES
InMotion Arm/MIT Manus	Bionik Labs	2D workspace with 2 DOF.	Three ^{13,22,23}
Kinarm Exoskeleton Robot Lab	Kinarm	2D workspace with 2 DOF.	Twelve ^{10,12,15-21,24,26,27}
PHANToM	3D Systems	3D workspace with 6 DOF.	Two ^{9,25}
REAPlan	Axinesis Rehabilitation Technologies	2D workspace with 2 DOF.	Two ^{11,14}

Abbreviations: 2/3D, 2/3-dimensional; DOF, Degrees of freedom.

The robotic measures of participant performance for point-to-point reaching tasks were also consistent among articles. These included at least 1 measure of path smoothness; however, the measures of smoothness differed among robotic configurations. The most common measures were the number of speed peaks (indicating a change in direction to correct for an error in movement)^{10,13,20,21,24,26} and different speed ratios,^{10,11,13,14,18,19,23} while normalized jerk of the movement was also used.^{13,23} Ideally these measures would be able to be combined and assessed in a meta-analysis, but due to differences in how the parameters were calculated, the data could not be combined.

Path following was also a common robotic task; however, the paradigms differed among robotic platforms. For example, with the NEPSY-II protocol,³⁵ a digitized version of a paper and pen task required participants to follow a curvilinear path,⁹ while other robotic systems had participants trace a circle and a square shown on the screen.^{14,23} The fundamental assessment and abilities being tested were the same between these 2 different paradigms, but the results could not be combined because of the differences in experimental setup and the methods by which measures were calculated.

Many other robotic tasks and measures were robot dependent and some were laboratory specific. For example, the free amplitude reaching task was only completed on the REAPlan robotic platform, and the test of kinesthesia was only completed using a Kinarm in a lab in Calgary, Alberta, Canada.^{10,14,18,19} While these tasks and measures were used in multiple studies, most of the overlap between populations and tasks were within specific research groups, not among groups.

Finally, the clinical measures used to assess different aspects of function were not standard among studies unless the studies were conducted by the same lab. Although this result is unsurprising when considering the number of potential clinical assessments that could be administered in typically developing populations, the limited overlap among studies that included the same pathological population is unexpected. For example, not all studies that included participants with CP recorded MACS level, which is a standardized self-assessment of upper-limb abilities.³⁴

Summaries of study aims, statistical tests used, results, and biases/limitations

Table 4 presents the aims, statistical tests and comparisons made, summary of results, and noted biases or limitations for each

study included in the review. The most common limitations were that sample sizes were not justified, and power analyses were not conducted. The most commonly noted bias was a sampling bias, as participants were often recruited from pre-existing research databases or through specific physicians.

Discussion

This systematic review aimed to determine what robotic tools and measures are being used to quantify upper-limb function in typically developing children and youth. Two-hundred and twenty-six articles were identified, 19 of which were included. Four robotic devices were used in the evaluations. Most of the articles (15 of 19) compared typically developing children to children with different pathologies, most commonly cerebral palsy. The most common robotic tests were point-to-point reaching tasks (without perturbations), and path following tasks. The most frequent measures of performance included movement smoothness, reaction time, and path length ratio (the ratio between a perfectly straight path between 2 points and the hand path of the participant).

Study populations

Study populations typically consisted of fewer than 50 participants per group, but 8 articles had more than 50 participants in at least 1 group. This size is similar to studies of the adult population as found by Sivan et al in their systematic review of robotic upper-limb therapy in people with stroke.² However, the sample sizes were typically much larger than in the Chen and Howard systematic review that evaluated upper-limb robotic therapy in children with CP.³ The increased sample sizes are promising as larger sample sizes lead to higher power of statistical results and more confidence in results and conclusions.

A study by Schönbrodt and Perugini used mathematical simulations to determine what sample size is needed to stabilize correlations. Their results indicated that sample sizes upwards of 250 datapoints are needed to stabilize a correlation in most cases.³⁶ Only 1 article included in this review had a sample size this large; however, that was the control group and the pathological population (children with CP) had only 3 participants.¹² The results from Schönbrodt and Perugini are important to consider for future research into the efficacy of robotic therapies. While studies with smaller sample sizes can

Table 3. Summary of upper-limb functions assessed, and robotic/clinical tests and measures.

ASPECT OF FUNCTION	ROBOTIC TESTS	ROBOTIC MEASURES	CLINICAL OUTCOME MEASURES (IF APPLICABLE)
Action planning	Target interception ⁹	Path length ratio ⁹	None ⁹
	Speed and path following ⁹	Movement duration ⁹	
	Path following (without perturbations) ⁹	Distance between gaze and end-effector position ⁹	
		Time on target ⁹	
		Normalized velocity ⁹	
Kinesthesia	Movement mirroring (speed and path mirroring) ^{10,18,19}	Response latency ^{10,18,19}	Upper-limb position sense ^{18,19}
		Path length ratio ^{10,18,19}	Thumb localization ^{18,19,28}
		Initial direction error ^{10,18,19}	Stereogenesis ^{18,19}
		Peak speed ratio ^{10,18,19}	Graphesthesia ^{18,19}
Motor adaptation/learning	Point to point reaching (with perturbations) ^{22,25}	Path deviation ^{22,25}	None ^{9,22,25}
	Path following (with perturbations) ⁹	Acceleration peak ²²	
	Object interception ⁹	Peak and average speed ²²	
	Speed and path following ⁹	Direct field effect ²⁵	
		Adaptation effect ^{22,25}	
		Adaptation rate ²⁵	
		Aftereffect ²⁵	
		Directional effects ²²	
		Path length ratio ⁹	
		Movement duration ⁹	
		Distance between gaze and end-effector position ⁹	
	Time on target ⁹		
Normalized velocity ⁹			
Movement smoothness (calculated as normalized jerk) ⁹			
Overall motor performance	Point to point reaching (without perturbations) ^{10-14,16,20,21,23,26}	Postural speed ^{10,20,21,26}	Assisting Hand Assessment (AHA) ^{15,16,19,20,21,29}
	Object interception (bimanual task) ^{10,16}	Reaction time ^{10,12,16,20,21,24,26}	Melbourne Assessment of Unilateral Upper Limb Function (MA) ^{15,16,19,20,21,22,30}
	Object interception and avoidance (bimanual task) ¹⁵	Initial direction error ^{10,13,16,20,21,26,24}	Chedoke-McMaster Stroke Assessment ^{20,21,31}
	Free amplitude reaching ¹⁴	Initial distance ratio ^{12,20,21,24,26}	Purdue Pegboard test (LaFayette Instrument Co, LaFayette, IN) ^{10,20,21}
	Path following (without perturbations) ^{14,23}	Initial speed ratio ^{12,20,21, 26}	Mirror-movement assessment ²¹

(Continued)

Table 3. (Continued)

ASPECT OF FUNCTION	ROBOTIC TESTS	ROBOTIC MEASURES	CLINICAL OUTCOME MEASURES (IF APPLICABLE)
		Movement smoothness (calculated as number of speed peaks ^{10,13,20,21,24,26} calculated as speed ratios, ^{11,13,14,23} calculated as normalized jerk ^{13,23})	Bruininks-Oseretsky Test of Motor Proficiency (Visual-Motor Control and Upper-Limb Speed and Dexterity subsets) ^{14,32}
		Minimum-maximum speed difference ^{20,21,24,26}	Scale for the Assessment and Rating of Ataxia ^{13,32,33}
		Movement time ^{10,12,13,20,21,24,26}	Manual Abilities Classification System (MACS) ^{11,12,17-21,34}
		Path length ratio ^{11,13,14,16,20,21,24,26}	Gross Motor Function Classification System
		Maximum speed ^{10,13,16,20,21,24,26}	
		Objects intercepted (total and with each hand) ^{10,15,16}	
		Median error ^{10,16}	
		Mean hand speed ^{10,11,13,16}	
		Hand movement area ¹⁶	
		Hand bias of interceptions ^{10,16}	
		Hand miss bias ¹⁶	
		Hand transition ¹⁶	(GMFCS) ¹²
		Hand movement area bias ^{10,16}	
		Hand speed bias ^{10,16}	
		No movement onset ²⁶	
		No movement end ²⁶	
		Reach amplitude ¹⁴	
		Movement speed ¹⁴	
		Reach accuracy ^{14,23}	
		Path deviation ^{13,14,23}	
		Duration of submovements ¹³	
		Amplitude of submovements ¹³	
		Distractor objects hit (total and with each hand) ¹⁵	
		Distractor proportion of hits ¹⁵	
		Object processing rate ¹⁵	
Position sense	Arm position mirroring ^{10,17,19,27}	Variability in mirroring ^{10,17,19,27}	Wrist and thumb position sense ^{17,19}
		Contraction/expansion ^{10,17,19}	Thumb localization ^{17,19}
		Systematic shift ^{10,17,19}	AHA ²⁷
			MA ²⁷
Strength	Isometric and isokinetic maximum voluntary contraction force measurements ¹¹	Isometric and isokinetic pushing forward and pulling backward force measurements ¹¹	Isometric maximum voluntary contraction measurements using a hand-held dynamometer ¹¹

Table 4. Description of aims of studies, statistical tests used, summary of results, and noted biases or limitations.

ARTICLE	AIMS OF STUDY	STATISTICAL TESTS USED	SUMMARY OF RESULTS	NOTED BIASES AND LIMITATIONS
Casellato et al, ⁹	To develop and validate an experimental setup for tracking and manipulating gaze and hand movements with a robotic pen apparatus. The performance of 1 participant with ASD was compared to a group of typically developing children.	Between group comparisons were made using the interquartile range (the 25th-75th percentile range) of performance of typically developing participants. Within group comparisons were made using the Mann-Whitney <i>U</i> -test.	The child with ASD showed intact, but delayed motor adaptation to an externally applied force field, and more variable gaze-hand patterns than typically developing children. The child with ASD also showed less efficient movements when trying to “catch” a target.	Limitations included: not justifying sample size, small sample size, and not performing a power analysis.
Cole et al, ¹⁰	To characterize sensorimotor learning changes caused by tDCS in typically developing children. To determine correlations between robotic and clinical measures (the PPB) of sensorimotor function. Comparisons were made between pre- and post-training with the PPB for groups who underwent sham tDCS, tDCS, and HD-tDCS	Between group comparisons were made using the Kruskal-Wallis 1-way ANOVA on ranks with Dunn’s test for post-hoc corrections. Within group comparisons were made with paired sample <i>t</i> -tests with Bonferroni’s post hoc corrections. Correlations between robotic and clinical measures were calculated with the Spearman’s rank order correlation.	Treatment with tDCS and HD-tDCS enhanced motor learning with a medium effect sizes as compared to sham stimulation. Reaching movements were faster and more accurate after training for all groups, but not different between groups. Mirror-matching of movements, a measure of kinesthesia, improved with training for all groups, but did not differ between groups. The HD-tDCS group showed significantly faster movements in an object hitting task compared to sham stimulation. Training and intervention did not change arm position matching performance. PPB score correlated with: reaction time (negative correlation), and total number of objects hit.	A recruitment bias was noted as some participants were recruited through a pre-existing research cohort. Limitations included: not justifying sample size, small sample size, and not performing a power analysis.
Dehem et al, ¹¹	To develop a robotic method to determine upper-limb working area, kinematics, and strength in typically developing children and children with CP. To compare kinematic measures between children with CP and typically developing children.	Relations between robotic and dynamometer measures of strength and computed using all participant data. Relations between kinematic indices, movement amplitude, and movement direction were computed in each group separately. Normality of the data and equality of variances were tested using the Shapiro-Wilk and Brown-Forsythe tests. Relations in normally distributed data were calculated with the Pearson’s correlation, and non-normally distributed data with the Spearman’s correlation. Force comparisons between groups were made with a <i>t</i> -test. Effects of movement direction on isokinetic force was analyzed with a 1-way ANOVA with Tukey’s post-hoc corrections.	Working area correlated to upper-limb length. Isometric force measured with the robot and hand-held dynamometer were correlated in both groups. Typically developing children generated larger forces than children with CP, and force generated was correlated to MACS level. Movement direction and force generated were correlated in each group. Kinematic measures were better in typically developing children than children with CP. Movement speed and straightness were positively correlated to movement amplitude, but no relation was found between smoothness and amplitude.	A recruitment bias was noted as typically developing children were drawn from a “convenience sample”, and children with CP were recruited through an existing CP reference centre. Limitations included: not justifying sample size, not performing a power analysis, and restricting participant age to 12-years-old.

(Continued)

Table 4. (Continued)

ARTICLE	AIMS OF STUDY	STATISTICAL TESTS USED	SUMMARY OF RESULTS	NOTED BIASES AND LIMITATIONS
Dobri et al. ¹²	To create normative models of performance on a point-to-point reaching task and assess if the models can differentiate between typical and impaired performance using data from children with CP.	Curves quadratic with age were fit to each robotic measurement parameter, and included dummy-coded (0 or 1) variables for participant sex and which hand (dominant or non-dominant) was performing the task. Models linear with age were used to account for changing variability in performance with age. The normative models were transformed into z-space and the 5th to 95th percentile ranges ($z +/ -1.64$) were used to differentiate typical from impaired performance.	The normative models differentiated performance of some of the participants with CP on some robotic measures.	A recruitment bias was noted as participants with CP were recruited through one physician. Limitations included: not justifying sample size, a small sample size of participants with CP, all participants with CP having the same MACS and GMFCS levels, not performing a power analysis.
Germanotta et al. ¹³	To assess the discriminative sensitivity of the robotic assessment between typically developing children and children with Friedreich's Ataxia, the reliability of selected indices that summarize aspects of motion from the robotic assessment, and to evaluate correlations between the robotic assessments and clinical scales.	Between group comparisons were completed with independent sample <i>t</i> -tests and Mann-Whitney <i>U</i> -tests. Correlations between robotic and clinical measures were computed using the Spearman's rank order correlation.	All but 2 robotic measures were significantly different between the control and pathological populations, and all indices were highly and reliably discriminative between the groups. Duration of movement, normalized jerk, and number of submovements were correlated to the progress of the ataxia, as measured using the SARA.	A recruitment bias was noted as participants with Friedreich's Ataxia were recruited from one location. Limitations included: not justifying sample size, small sample size, and not performing a power analysis.
Gilliaux et al. ¹⁴	To assess age effects on upper-limb function and develop normative models in typically developing children, and establish construct validity for the tests and robotic apparatus used. Construct validity was tested by computing correlations between the robotic measures of performance and the BOTMP Visual-Motor Control and Upper-Limb Speed and Dexterity subsets.	Regression analysis using exponential models. Spearman's rank order correlation between robotic measures and the BOTMP Visual-Motor Control and Upper-Limb Speed and Dexterity subsets.	Established age-related normative models for a total of 28 measuring parameters for 4 different motor control tasks. Significant correlations between robotic measures and the BOTMP were found in 25 measures, and 24 measures showed significant age effects ($ r > 0.3$).	No biases were noted. Limitations included: not justifying sample size, not performing a power analysis, restricting participant age to 12-years-old, and that the BOTMP shows ceiling effects as early as 8-years-old.

(Continued)

Table 4. (Continued)

ARTICLE	AIMS OF STUDY	STATISTICAL TESTS USED	SUMMARY OF RESULTS	NOTED BIASES AND LIMITATIONS
Hawe et al, ¹⁵	<p>To use an object hit and avoid task to determine impairments in children with hemiparetic CP to make rapid motor decisions compared to typically developing children, examine age effects on these abilities, and compare differences between the object hit and avoid task and a simpler, object hit version of the task.</p>	<p>Normality and homogeneity of variances were tested with the Shapiro-Wilk's test (and visual inspection) and Levene's test. Between group comparisons were made with ANOVAs, accounting for age. Post-hoc <i>t</i>-tests with Bonferroni corrections were used as well. Regression with a second order polynomial was used to create normative ranges to identify individual impairment based on age. Impairment was performance outside the 95% range. Cohen's <i>d</i> was used to calculate effect size between the object hit and avoid, and object hit task. Pearson correlations were used to calculate correlations between robotic and clinical measures.</p>	<p>Significant differences were found between the typically developing group and AIS group for all 7 robotic parameters, between the typically developing group and PVI group for 6 parameters, and between the AIS and PVI group for 2 parameters. Complexity of task (object hit and avoid or object hit only) affected percentage of targets hit with the dominant arm in typically developing participants, and number of objects hit with either the dominant or non-dominant arm for children with CP. The Melbourne Assessment and Assisting Hand Assessment correlated to measures of targets hit.</p>	<p>A recruitment bias was noted as participants were recruited through a pre-existing research cohort and robotics stroke lab. Limitations included: Not justifying sample size, not performing a power analysis, and performing tests which require normally distributed data after confirming the data were not normally distributed. The data failed the Shapiro-Wilk's test, but passed "visual inspection".</p>
Hawe et al, ¹⁶	<p>To quantify differences in motor function between children with CP and typically developing children using a robotic object hitting task, compare performance of the object hitting task to a point-to-point reaching task, and look at relationships between the robotic and clinical measures of motor and visuospatial attention.</p>	<p>Between group comparisons were made with ANCOVAs with Bonferroni post-hoc corrections. Regression was used to determine relationships between object hit task parameters and point-to-point reaching parameters. Regression with a second order polynomial was used to create normative ranges to identify individual impairment based on age. Impairment was performance outside the 95% range.</p>	<p>Overall performance was significantly poorer in children with CP compared to controls. Significant differences were noted for some measures of performance between the 2 groups with CP. Hand speed of the affected hand of participants with CP was significantly slower than typically developing children, and movement area for the affected arm of children with AIS was significantly different than typically developing children. Hand speed and movement area accounted for variance in number of objects hit, compared between groups. Reaction time, initial direction error, and path length ratio from the reaching task were related to number of objects hit. Clinical measures were related to number of objects hit.</p>	<p>A recruitment bias was noted as participants were recruited through a pre-existing research cohort and robotics stroke lab. Limitations included: Not justifying sample size, not performing a power analysis, and some statistical tests required Gaussian distributions in the data but the authors did not state if the distribution was assessed.</p>
Kuczynski et al, ¹⁷	<p>To quantify impairment of position sense in children with cerebral palsy compared to typically developing children, and to determine correlations between robotic and clinical measures of position sense.</p>	<p>Between group comparisons were made with ANOVAs with Tukey's post-hoc corrections. Within group comparisons were made with paired sample <i>t</i>-tests and the Wilcoxon signed rank sum test. Correlations between robotic and clinical measures were calculated using the Spearman's rank order correlation, and ANOVAs.</p>	<p>Variability in position matching and spatial shift were significant between the control group and the children with cerebral palsy, but the other robotic measure was not. AHA, and PPB were correlated to variability and spatial shift in children with AIS, but not the control group or children with PVI. Impaired thumb localization, stereogenesis, and graphesthesia were associated with variability in children with AIS.</p>	<p>A recruitment bias was noted as participants were recruited through a pre-existing research cohort. Limitations included: a small sample size in each subgroup of children with cerebral palsy.</p>

(Continued)

Table 4. (Continued)

ARTICLE	AIMS OF STUDY	STATISTICAL TESTS USED	SUMMARY OF RESULTS	NOTED BIASES AND LIMITATIONS
Kuczynski et al. ¹⁸	To quantify kinesthetic deficits in children with hemiparetic perinatal stroke compared to typically developing children, and determine correlations between robotic and clinical measures of kinesthesia.	Between group comparisons were made with ANCOVAs with Bonferroni post-hoc corrections, independent sample <i>t</i> -tests, and the Mann-Whitney <i>U</i> -test. Within group comparisons were made with paired sample <i>t</i> -tests and the Mann-Whitney <i>U</i> -test. Linear regression models assessed effects of age on robotic measures of kinesthesia. Correlations between robotic and clinical measures were assessed with the Spearman's rank order correlation, the Pearson Correlation, and chi-squared test.	Participants with perinatal stroke exhibited significantly impaired kinesthesia compared to controls. Significantly different measures included response latency, initial direction error, path length ratio, and variability within each robotic measures of kinesthesia. Significant differences between stroke type were found in 2 of 8 robotic measures. The AHA, CMSA, and clinical measures of position sense did not correlate with robotic measures of kinesthesia. Clinical measures of stereogenesis, graphesthesia, and the Purdue Pegboard did correlate with robotic measures of kinesthesia.	A recruitment bias was noted as participants were recruited through a pre-existing research cohort. Limitations included: not justifying sample size, and not performing a power analysis.
Kuczynski et al. ¹⁹	To characterize the relationship between structural connectivity in the dorsal column medial lemniscus (DCML) and proprioceptive dysfunction in children with AIS and PVI. Comparisons were made between structural connectivity and robotic measures of proprioceptive function within and between groups of children with AIS, PVI, and typically developing children. Correlations between robotic and clinical measures of proprioception were also made.	Between group comparisons were made with ANCOVAs with Bonferroni post-hoc corrections. Within group comparisons were made with paired sample <i>t</i> -tests. Correlations between robotic and clinical measures of proprioception were made using Partial Spearman's correlations. Partial Spearman's correlations were also used to study the relationship between robotic measures and structural connectivity in the DCML.	Participants with AIS and PVI both showed correlations between impaired proprioception and lesioned hemisphere DCML tract diffusion properties. Lesioned hemisphere tract diffusion was significantly different between typically developing children and children with AIS and PVI, but no difference was observed in contralesional hemispheres. Participants with AIS and PVI showed significant differences in variability in arm position matching, response latency and its variability, initial direction error and its variability, peak speed ratio variability, and path length ratio variability in robotic tests of kinesthesia as compared to typically developing children. Participants with AIS also showed significant differences in path length ratio as compared to controls.	A recruitment bias was noted as participants were recruited through a pre-existing research cohort. Limitations included: not justifying sample size, small sample size, and not performing a power analysis.
Kuczynski et al. ²⁰	To quantify the kinematics of reaching in both arms of hemiparetic children with perinatal stroke, compare kinematic reaching between typically developing children and 2 different populations of children with differing diagnoses of perinatal stroke (AIS and PVI), and to assess correlation between the robotic kinematic reaching parameters and clinical motor assessments.	Between group comparisons were made using ANCOVAs with Bonferroni post-hoc corrections. Within group comparisons were made with paired-sample <i>t</i> -tests and the Mann-Whitney <i>U</i> -test. Partial Spearman's correlation, controlling for age, were used for correlations between clinical tests and robotic measures.	Significant differences were found between control groups and both groups of children with perinatal stroke, regardless of reaching arm (contralesional or ipsilesional arm). The group with AIS also showed significant deficits compared to the PVI group. Reaching parameters with the contralesional arm showed "modest" correlation with clinical measures in the AIS group only.	A recruitment bias was noted as participants were recruited through a pre-existing research cohort and robotics stroke lab. Limitations included: not justifying sample size, and not performing a power analysis.

(Continued)

Table 4. (Continued)

ARTICLE	AIMS OF STUDY	STATISTICAL TESTS USED	SUMMARY OF RESULTS	NOTED BIASES AND LIMITATIONS
Kuczynski et al, ²¹	To characterize differences in CST structural connectivity between hemispheres in children with perinatal stroke and typically developing children, and to determine relationships between the connectivity and motor function using clinical and robotic measures.	Between group comparisons in motor function and diffusion parameters in each hemisphere were made using ANCOVAs with Bonferroni post-hoc corrections. Paired sample <i>t</i> -tests were used for within group comparisons. Linear age-models were used to characterize developmental trajectories of diffusion tensor imaging (DTI) of the CSTs. Relationships between DTI, robotic measures, and clinical measures were calculated with partial Spearman's correlations, controlling for age.	All measures of CST diffusion in the lesioned hemispheres differed between the control and AIS group, while only one parameter differed between controls and the PVI group. Non-lesioned hemispheres did not show significant differences from the controls. Lesioned CST parameters correlated with robotic measures of reaching movements of the affected limb. Few correlations were found between the non-lesioned CST parameters and unaffected arm reaching movements.	A recruitment bias was noted as participants were recruited through a pre-existing research cohort. Limitations included: not justifying sample size, not performing a power analysis, small sample size, and some statistical tests required Gaussian distributions in the data but the authors did not state if the distribution was assessed.
Masia et al, ²²	To assess to what extent motor adaptation to an external force field is impaired in children with CP.	Between group and within group comparisons were completed with ANOVAs.	Both control groups and participants with CP showed motor adaptation, but the adaptation rate was not significant, and the learning index was lower in participants with CP.	No biases noted. Limitations included: not justifying sample size, small sample size, not performing a power analysis, and some statistical tests required Gaussian distributions in the data but the authors did not state if the distribution was assessed.
Pacilli et al, ²³	To assess whether control groups need to be age matched to pathological populations for robot-mediated therapy clinical studies involving pediatric populations. This was tested by assessing if movement coordination, accuracy, and smoothness for point-to-point reaching and circle drawing tasks differ between child and adult populations.	Between group comparisons were made with ANOVAs. Within group comparisons were made with paired sample <i>t</i> -tests, and ANOVAs with Greenhouse-Geisser post-hoc corrections.	Smoothness of the circle drawing task was significantly different between children and adults, but accuracy and joint coordination were not. Children performed the reaching task less smoothly and accurately than adults. The results indicate age-matched controls are necessary for future work with robot mediated therapies.	A potential recruitment bias existed as a "convenience sample" was taken for participants in the study. Limitations included: not justifying sample size, small sample size, and not performing a power analysis.

(Continued)

Table 4. (Continued)

ARTICLE	AIMS OF STUDY	STATISTICAL TESTS USED	SUMMARY OF RESULTS	NOTED BIASES AND LIMITATIONS
Skarsgard et al. ²⁴	To work towards identifying DCD using exploratory factor analysis (EFA) and by creating normative models of typical performance with a robotic point-to-point reaching task.	EFA and normative models of performance were used to differentiate participants with DCD from typically developing children. Normative models were created using multivariable regression to account for non-linear changes with age, and binary differences between participant sex and hand performing the task (dominant vs non-dominant hand).	Both EFA and the normative models showed ability to differentiate between participants with DCD and typically developing children.	A recruitment bias was noted as participants with DCD were recruited through one physician. Limitations included: not stating what type of non-linear regression model was used, not justifying sample size, a small sample size of participants with DCD, and not performing a power analysis.
Takahashi et al. ²⁵	To test the hypothesis that children use the same adaptive mechanisms as adults to adapt to environmental noise. The hypothesis was tested by comparing motor performance of children and adults during reaching movements after adapting to variable robot-applied force fields.	Between group comparisons were made with ANOVAs with Bonferroni post-hoc corrections. Within group comparisons were made with paired sample <i>t</i> -tests. Linear regression was used to determine age-effects on robotic measures of task performance.	Children were able to adapt to force fields at a similar level as adults, indicating that the ability to adapt to environmental noise is fully developed in childhood. Performance was still more variable in children than adults after adaptation, indicating that movement inconsistency limits motor performance rather than adaptation ability.	No biases were noted. Limitations included: not justifying sample size, small sample size, and some statistical tests required Gaussian distributions in the data but the authors did not state if the distribution was assessed.
Williams et al. ²⁶	To quantify sensorimotor impairments in children with FASD compared to controls using a robotic point-to-point reaching task.	Between group comparisons were made with ANOVAs and multivariate ANOVAs. Group effect size was calculated using Cohen's <i>d</i> . Linear regression with age was computed to quantify typical performance and create normative ranges of performance to identify impaired typically developing participant performance (performance outside 95% of performance, either 1 or 2-sided).	Group effect sizes were small for reaction time, movement time, and no movement endpoints. Hand speed ratio had a medium effect size. Large effects ($d > 0.8$) were observed for postural stability, initial direction error, initial movement ratio, number of speed peaks, min/max speed difference, and path length ratio. Children with FASD experience significantly impaired reaching compared to typically developing children.	Limitations included: not justifying sample size, not performing a power analysis, and splitting typically developing participants into smaller groups for some parameters (which reduced power of the results).
Woodward et al. ²⁷	To better understand the relationship between sensorimotor functional connectivity in the brain and sensorimotor deficits in children with CP compared to typically developing children.	Data were tested for normality with the Shapiro-Wilk's test. Between group comparisons were made with the Student <i>t</i> -test. Pearson correlations were used to quantify between-node connectivity strength, age, gender, and sensorimotor function.	Increased connectivity between brain nodes was correlated to less variability in a position matching task for participants with CP. The correlation was the opposite (higher connectivity lead to more variability) in typically developing children. AHA scores were negatively correlated to connectivity in participants with CP.	A recruitment bias was noted as participants were recruited through a pre-existing research cohort. Limitations included: not justifying sample size, small sample size, and not performing a power analysis.

Abbreviations: CP, Cerebral Palsy; AIS, Arterial Ischemic Stroke; PVI, Periventricular Infarction; MACS, Manual Ability Classification System; ANOVA, Analysis of Variance; ANCOVA, Analysis of Covariance; BOTMP, Bruininks-Oseretsky Test of Motor Performance; CST, corticospinal tract; AHA, Assisting Hand Assessment; CMSA, Chedoke-McMaster Stroke Assessment; DCD, Developmental Coordination Disorder; FASD, Fetal Alcohol Spectrum Disorder; ASD, Autism Spectrum Disorder; PPB, Perdue Pegboard; (HD-)IDCS, (High Definition-)transcranial direct-current stimulation.

give an indication of the efficacy of robotic therapies compared to traditional methods, the true difference will only be known once sample sizes reach 250 participants or more.

While only 2 articles stated age- and/or sex-matched control populations, all research articles stated that the age demographics of the control group were similar to the pathological population. The similarities in age ranges are important for work with children and youth as motor performance is known to improve significantly with age.³⁷ Matching age will be important for future research comparing robotic therapies to traditional therapy methods to ensure confounding factors, such as age-related performance differences, are minimized. The article that tested differences in male and female performance²⁶ did not find significant differences, which is consistent with other previously published data. These results indicate that it may not be necessary to have sex-matched controls, or sex-matched intervention groups.^{37,38}

Robotic assessments

Only 1 robotic device could assess both arms simultaneously, and 8 of the articles assessed both hands of each participant. Two of the articles that tested and compared performance between the dominant and non-dominant arms of participants found significant differences. These results indicate that it may be important to separate dominant and non-dominant arms in analysis of unimanual tasks and to ensure that the hand dominance is considered when comparing between groups. The studies did not compare right versus left dominance, only dominant versus non-dominant arms.

Point-to-point reaching tasks (without perturbations) are common assessments of motor function as goal-directed reaching is used in many aspects of daily living. These types of tasks are included in different clinical assessments such as the Melbourne Assessment of Unilateral Upper Limb Function, Perdue Pegboard (PPB) test, and the Bruininks-Oseretsky Test of Motor Proficiency.^{30,32,39} By adapting point-to-point tasks to robotic platforms, more precision and objectivity can be attained.

Point-to-point reaching tasks occurred on all 4 robotic platforms and were typically used to quantify overall motor function. These robotic tasks are similar in nature to accepted clinical assessments of motor function. For example, the PPB consists of goal-directed point-to-point reaching.³⁹ Speed and accuracy of movement are important and emphasized in both the robotic point-to-point reaching tasks and the PPB. Robotic tasks and the PPB quantify different aspects of the reaching movement. The PPB gives aggregate measures of function (the number of pegs placed, and assemblies built), which is also affected by manual dexterity, and can identify impairment but not what part of the movement is impaired. Most of the robotic tasks do not measure manual dexterity but can quantify and differentiate aspects of reaching movements to give precise information on motor control differences identifying

impairment. For example, robotic measures include the path length ratio (the ratio between the path followed by the participant and the shortest line between the 2 points), path smoothness, and total movement time. These 3 measures give an idea of different ways the point-to-point reaching may be affected, rather than simply identifying that it is impaired. A clinician administering the PPB may be able to make qualitative assessments, but there is no quantification of which part of the movement is most impaired. The Melbourne Assessment and Bruininks-Oseretsky Test have more specific outcome measures than the PPB that can help identify aspects of movements that are most affected by a pathology.

The addition of perturbations to point-to-point reaching and path following tasks allow robotic assessments to evaluate aspects of function that may not be easily tested with existing clinical tests. Specifically, motor learning/adaptation to perturbations could not be directly tested with the clinical tests used by any of the included studies; however, robots could apply perturbations to assess how participants adapt and learn to compensate for external perturbations. These types of tests allow researchers to determine different mechanisms for how a pathology may affect a person's motor function. They may also be clinically relevant as the tasks can measure how well someone can recover from an unexpected perturbation (such as an object shifting when it is being carried), or can be included in therapies to improve someone's ability to adapt to unexpected perturbations.

While the precision and specificity of robotic measures can be helpful, the number of measures taken from 1 test can be overwhelming, and not all measures may be clinically relevant. For example, the object interception and avoidance task used in Hawe et al has 17 different performance measures,¹⁵ and the point-to-point reaching task on the same robotic platform has 11 performance measures.^{12,26} The Melbourne Assessment has 16 tasks in the entire assessment,³⁰ similar to the object interception and avoidance task, but many of the articles used multiple robotic tasks for similar evaluations. This increased the number of performance measures available with the robot as compared to a single clinical test. Additionally, clinical tests verify that all sub-tasks and measures are clinically relevant, whereas the robotic tasks and measures have not been tested to the same degree for clinical significance. Finally, these clinical tests can be compared to existing normative ranges to identify impairment with overall scores. Many of the robotic tasks do not have normative ranges of performance against which to compare. Those that do include normative ranges are confined to each specific measure rather than overall measures of performance making comparison to impaired populations more difficult. As noted by Dobri et al, individual measures of performance from 1 task are not sufficient to identify impairment.¹² If robots are to be used to identify impairment of upper-limb motor function in children and youth, normative ranges across platforms are needed as well as aggregate measures of performance for each task in a similar manner to what exists for clinical tests.

Many robotic measures did not correlate with clinical measures of motor performance. The lack of correlation could be due to differences in the actual measure between the robotic and clinical tests, as well as measurement limitations of each tool.¹⁴

Data analysis

Many of the statistical tests require specific criteria to be met for them to be valid. For example, *t*-tests, ANOVAs, and ANCOVAs require data distributions to be Gaussian and equal variance between datasets. Seventeen articles completed tests that required Gaussian distributions of data^{10-13,15-27} whereas 13 stated that the distribution was evaluated.^{10-13,15,18-20,23,24,26,27} One article ignored the results of the test for a Gaussian distribution as the data appeared to follow this distribution upon “visual inspection.”¹⁵ Seven of these articles either used a different test which does not require a Gaussian distribution or applied transformations to make the distribution Gaussian.^{9,13,17,18,20,24,26} Only 1 article stated explicitly that methods were undertaken that corrected unequal variance between datasets.¹³

Limitations of research

Many of the included articles did not explicitly state the necessary conditions for statistical analyses. It is therefore unknown if the results of these analyses are accurate.

While many articles have large study populations (more than 50 participants in at least 1 group), most articles were limited by small sample size in at least 1 group. The articles with small populations did report the sample size as a limitation and stated the results were a proof of concept, rather than an in-depth analysis.

Limitations of review

One limitation of this review is that only 1 reviewer screened article titles and abstracts. Typically, this process is completed by at least 2 reviewers to reduce potential biases in article selection. Another limitation was that a meta-analysis was not completed due to the non-homogeneity of the study paradigms and robotic measures. While a meta-analysis could be completed on tasks that were used in multiple studies, repetition of participant data was included in several articles from the same lab; analysis that included the same participants more than once would skew the analysis.

Recommendations

In future studies and robot design, the same robotic tasks and measures should be used to quantify motor function. Increased homogeneity of study methods would enable meta-analyses to be effectively conducted to increase the statistical power of the findings. The authors recognize that this recommendation may be difficult to achieve as the robotic apparatuses are all commercially

available, and cooperation between competitors to ensure consistency of tasks and measures across platforms is unlikely. As the popularity of these devices grow, the diversity of groups using each apparatus may accommodate future meta-analyses.

Similarly, it is recommended that future studies compare the same clinical tools and measures to assess motor function to facilitate comparisons among studies. The use of the same clinical measures would be easier to achieve than the same robotic tasks and measures as the different measuring tools generally are not directly competitive and are significantly cheaper than robots. The lower cost would allow researchers to have access to multiple clinical assessment tools to ensure their findings can be compared to other research in the same area.

Additionally, future researchers should consistently report testing hand/handedness, justify sample size and compute power analyses, and state whether requirements for statistical tests have been met. Testing hand/handedness is important to report as it allows other researchers to fully reproduce experiments and make more complete comparisons between previous work and their own. It is also important to justify sample size and compute power analyses to better understand the statistical strength of the study results. Finally, clearly stating if requirements for statistical tests have been met increases the confidence other researchers have in the results of the study.

Contributions of review and implications for practice and research

Unlike previous research, this review has compiled and summarized robotic devices, tasks, and outcome measures that have been used to quantify upper-limb motor function in typically developing children and youth. It further summarizes those articles whose data were compared to pathological populations and related clinical measures. This compilation will allow clinicians and researchers to quickly and easily identify the devices, tasks, and measures currently available to quantify upper-limb function. Clinicians can now make educated decisions about task type and outcome measures that may be most useful for patient measurement. Researchers can more easily identify methods to ensure effective comparison across research groups as well as identify gaps in knowledge with respect to upper-limb function.

This review identified the need for normative databases of robotic tasks and outcome measures to enable the identification of impairment in individual children and youth, similar to clinical tests. The lack of normative databases prohibits impairment from being identified by clinicians without first collecting and analyzing their own database of results from typically developing children and youth. To make robotic devices usable and effective for clinicians, researchers will need to create these databases and make them readily available. There is precedent for this, as some robots do have normative databases for adults that are used to identify measures outside the typical performance ranges.⁴⁰

Conclusion

This systematic review was conducted to identify robotic tools and tests used in published literature to quantify upper-limb function in typically developing children. It also related these tools and techniques to outcome measures used for comparison to clinical populations. Fifteen of the 19 articles studied both typically developing and pathological populations, while 4 quantified motor function in only typically developing populations. While there was some overlap among studies in terms of population, robotic apparatus and tasks, and clinical measures of motor function, there was insufficient overlap to conduct meta-analyses. Future work should aim to use the same robotic apparatuses, tasks, and clinical measures to allow meta-analyses to be conducted to increase the statistical strength of the overall findings.

Author Contributions

SCD Dobri was involved in all stages of the production of the review and was the reviewer who screened titles and abstracts. HM Ready was involved in assessing full articles, writing, and editing the review. TC Davies was involved in developing the research questions and search string, solving disagreements on article inclusion at the full-text stage, and writing and editing the review.

ORCID iDs

Stephan CD Dobri  <https://orcid.org/0000-0002-8465-863X>

Theresa Claire Davies  <https://orcid.org/0000-0003-4880-2654>

REFERENCES

- Scott SH, Dukelow SP. Potential of robots as next-generation technology for clinical assessment of neurological disorders and upper-limb therapy. *J Rehabil Res Dev*. 2011;48:335.
- Sivan M, O'Connor R, Makower S, Levesley M, Bhakta B. Systematic review of outcome measures used in the evaluation of robot-assisted upper limb exercise in stroke. *J Rehabil Med*. 2011;43:181-189.
- Chen Y-P, Howard AM. Effects of robotic therapy on upper-extremity function in children with cerebral palsy: a systematic review. *Dev Neurorehabil*. 2016;19:64-71.
- Whitten TA, Mang CS, Cosh MS, Scott SH, Dukelow SP, Benson BW. Spatial working memory performance following acute sport-related concussion. *J Concussion*. 2018;2:2059700218797818.
- Lowrey C, Jackson CPT, Dukelow SP, Scott SH. A novel robotic task for assessing impairments in bimanual coordination post-stroke. *Int J Phys Med Rehabil*. 2014;53.
- Dukelow SP, Herter TM, Moore KD, et al. Quantitative assessment of limb position sense following stroke. *Neurorehabil Neural Repair*. 2010;24:178-187.
- Coderre AM, Amr Abou Zeid AA, Dukelow SP, et al. Assessment of upper-limb sensorimotor function of subacute stroke patients using visually guided reaching. *Neurorehabil Neural Repair*. 2010;24:528-541.
- Gilliaux M, Lejeune TM, Sapin J, Dehez B, Stoquart G, Detrembleur C. Age effects on upper limb kinematics assessed by the REAplan robot in healthy subjects aged 3 to 93 years. *Anna Biomed Eng*. 2016;44:1224-1233.
- Casellato C, Gandolla M, Crippa A, Pedrocchi A. Robotic set-up to quantify hand-eye behavior in motor execution and learning of children with autism spectrum disorder. In: *2017 International Conference on Rehabilitation Robotics*. IEEE; 2017: 953-958.
- Cole L, Dukelow SP, Giuffre A, Nettel-Aguirre A, Metzler MJ, Kirton A. Sensorimotor robotic measures of tDCS- and HD-tDCS-enhanced motor learning in children. *Neural Plast*. 2018;2018:5317405.
- Dehem S, Montedoro V, Brouwers I, et al. Validation of a robot serious game assessment protocol for upper limb motor impairment in children with cerebral palsy. *NeuroRehabilitation*. 2019;45:137-149.
- Dobri SCD, Samdup D, Scott SH, Davies TC. Differentiating motor coordination and position sense in children with Cerebral Palsy and typically developing populations through robotic assessments*. Paper presented at: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); July 20-24, 2020; Montreal, QC.
- Germanotta M, Vasco G, Petrarca M, et al. Robotic and clinical evaluation of upper limb motor performance in patients with Friedreich's Ataxia: an observational study. *J Neuroeng Rehabil*. 2015;12:41.
- Gilliaux M, Dierckx F, Vanden Berghe L, et al. Age effects on upper limb kinematics assessed by the REAplan robot in healthy school-aged children. *Ann Biomed Eng*. 2015;43:1123-1131.
- Hawe RL, Kuczynski AM, Kirton A, Dukelow SP. Robotic assessment of rapid motor decision making in children with perinatal stroke. *J Neuroeng Rehabil*. 2020;17:94.
- Hawe RL, Kuczynski AM, Kirton A, Dukelow SP. Assessment of bilateral motor skills and visuospatial attention in children with perinatal stroke using a robotic object hitting task. *J Neuroeng Rehabil*. 2020;17:18.
- Kuczynski AM, Dukelow SP, Semrau JA, Kirton A. Robotic quantification of position sense in children with perinatal stroke. *Neurorehabil Neural Repair*. 2016;30:762-772.
- Kuczynski AM, Semrau JA, Kirton A, Dukelow SP. Kinesthetic deficits after perinatal stroke: robotic measurement in hemiparetic children. *J Neuroeng Rehabil*. 2017;14:13.
- Kuczynski AM, Carlson HL, Lebel C, et al. Sensory tractography and robot-quantified proprioception in hemiparetic children with perinatal stroke. *Hum Brain Map*. 2017;38:2424-2440.
- Kuczynski AM, Kirton A, Semrau JA, Dukelow SP. Bilateral reaching deficits after unilateral perinatal ischemic stroke: a population-based case-control study. *J Neuroeng Rehabil*. 2018;15:77.
- Kuczynski AM, Dukelow SP, Hodge JA, et al. Corticospinal tract diffusion properties and robotic visually guided reaching in children with hemiparetic cerebral palsy. *Hum Brain Map*. 2018;39:1130-1144.
- Masia L, Frascarelli F, Morasso P, et al. Reduced short term adaptation to robot generated dynamic environment in children affected by Cerebral Palsy. *J Neuroeng Rehabil*. 2011;8:28, 12.
- Pacilli A, Germanotta M, Rossi S, Cappa P. Quantification of age-related differences in reaching and circle-drawing using a robotic rehabilitation device. *Appl Bionics Biomech*. 2014;11:91-104.
- Skarsgard M, Dobri SCD, Samdup D, Scott SH, Davies TC. Toward robot-assisted diagnosis of developmental coordination disorder. *IEEE Robot Autom Lett*. 2019;4:346-350.
- Takahashi CD, Nemet D, Rose-Gottron CM, Larson JK, Cooper DM, Reinkensmeyer DJ. Neuromotor noise limits motor performance, but not motor adaptation, in children. *J Neurophysiol*. 2003;90:703-711.
- Williams L, Jackson CPT, Choe N, Pelland L, Scott SH, Reynolds JN. Sensory-motor deficits in children with fetal alcohol spectrum disorder assessed using a robotic virtual reality platform. *Alcohol Clin Exp Res*. 2014;38:116-125.
- Woodward KE, Carlson HL, Kuczynski A, Saunders J, Hodge J, Kirton A. Sensory-motor network functional connectivity in children with unilateral cerebral palsy secondary to perinatal stroke. *Neuroimage Clin*. 2019;21:101670.
- Hirayama K, Fukutake T, Kawamura M. 'Thumb localizing test' for detecting a lesion in the posterior column-medial lemniscal system. *J Neurol Sci*. 1999;167:45-49.
- Krumlinde-Sundholm L, Eliasson A-C. Development of the Assisting Hand Assessment: a Rasch-built measure intended for children with unilateral upper limb impairments. *Scand J Occup Ther*. 2003;10:16-26.
- Bourke-Taylor H. Melbourne assessment of unilateral upper limb function: construct validity and correlation with the pediatric evaluation of disability inventory. *Dev Med Child Neurol*. 2003;45:92-96.
- Barreca SR, Stratford PW, Lambert CL, Masters LM, Streiner DL. Test-retest reliability, validity, and sensitivity of the Chedoke arm and hand activity inventory: a new measure of upper-limb function for survivors of stroke. *Arch Phys Med Rehabil*. 2005;86:1616-1622.
- Bruininks RH. *Bruininks-Oseretsky Test of Motor Proficiency*. American Guidance Service Circle Pines; 1978.
- Schmitz-Hübisch T, du Montcel ST, Baliko L, et al. Scale for the assessment and rating of ataxia. *Neurology*. 2006;66:1717.
- Eliasson A-C, Krumlinde-Sundholm L, Rösblad B, et al. The manual ability classification system (MACS) for children with cerebral palsy: scale development and evidence of validity and reliability. *Dev Med Child Neurol*. 2006;48:549-554.
- Brooks BL, Sherman EMS, Strauss E. NEPSY-II: a developmental neuropsychological assessment, Second Edition. *Child Neuropsychol*. 2009;16:80-101.
- Schönbrodt FD, Perugini M. At what sample size do correlations stabilize? *J Res Pers*. 2013;47:609-612.
- Payne VG, Isaacs LD. *Human Motor Development: A Lifespan Approach*. Routledge; 2017.
- Thomas JR, French KE. Gender differences across age in motor performance: a meta-analysis. *Psychol Bull*. 1985;98:260.
- Tiffin J, Asher EJ. The Purdue Pegboard: norms and studies of reliability and validity. *J Appl Psychol*. 1948;32:234-247.
- Dexterit-E 3.6 User Guide, Kinarm, Kingston, Ontario, Canada; 2016.