Preferred Running Head: Outcome Measures Commonly Used in JIA

Title: Psychometric Considerations and Age-Appropriateness of Outcome Measures Implemented in Exercise Randomized Controlled Trials Within the Juvenile Idiopathic Arthritis Cohort: A Systematic Review

Manuscript Word Count: 4,803

Abstract

Background: Juvenile idiopathic arthritis (JIA) is an autoimmune condition of multifactorial etiology resulting in chronic inflammatory joint disease, which may be associated with systemic manifestations. Therapeutic exercise is essential to counteract physical impairments, which requires the implementation of outcome measures (OMs) in research and practice, as they provide meaningful results for research efficacy, exercise program evaluation and quality, medication tolerance, and patient improvement. ***Purpose:*** To assess the types of OMs implemented in exercise randomized controlled trials (RCTs) related to the JIA cohort and the psychometric properties and age-appropriateness of the implemented OMs. ***Methods:*** The review was registered with PROSPERO (CRD42022336345) on May 30, 2022, followed by a systematic search across PubMed, EBSCOhost, Web of Science, and Ovid. Studies included were appraised using the Joanna Briggs Critical Appraisal Tool for RCTs. All data collection occurred according to the Preferred Reporting Items for Systematic Reviews and Meta-analysis. ***Results**:* A total of 51 outcome measures were implemented across the 20 RCTs: two clinician-reported OMs (4%), 19 patient or parent-reported OMs (37%), and 30 physical performance OMs (59%). The vast majority of included OMs increases the difficulty of comparison across studies and indicates a lack of consideration for validity, reliability, and age-appropriateness.

Keywords**:** physical health status, outcome measures, juvenile idiopathic arthritis, exercise, rehabilitation.

Introduction

Juvenile idiopathic arthritis (JIA) is a chronic autoimmune condition resulting in an inflammatory joint disease.[1] The condition is considered the most widespread chronic rheumatologic disease in children. The global prevalence of JIA is 3.8-400/100,000 children and an incidence of 1.6-23/100,000.[2–4] It encompasses a group of heterogenic inflammatory joint diseases of unknown etiology that occur before 16 years of age and last six weeks or more.[5]

Juvenile idiopathic arthritis can result in short- and long-term disability, impacting aerobic fitness, muscle strength, bone density, range of motion (ROM), physical functioning, impaired proprioception, and quality of life (QoL).[2] Children tend to become less physically active due to the disease impacting numerous organ systems.[2] However, Kuntze and colleagues conducted a systematic review that showed that physical activity could be well-tolerated and safe for children with JIA.[2,6,7] In addition, the study showed that improvements in balance, muscle strength, functional capacity, and QoL could be made through physical activity (PA) in children with JIA.[7] More specifically, Klepper reported that an exercise program of 30 to 50 minutes, two to three times a week, for 12 to 24 weeks can improve ROM, knee strength, functional capability, pain, and QoL.[2]

Research and practice must implement outcome measures (OMs) to appropriately assess an exercise intervention's efficacy and determine exercise prescription. Outcome measures are used in domains of performance, functionality, and participation.[8,9] In treatment (rehabilitation or medication), OMs also provide meaningful results for research, exercise program evaluation and quality, medication tolerance, patient improvement, and case management.[8–10] Specifically, in exercise therapy, OMs provide baseline measurements, a method to monitor patient and treatment progress, and to determine whether the final exercise outcome has been met.[8,11] Physicians apply OMs similarly to monitor and manage the disease through medication. In therapeutic exercise therapy, OMs form the essential core of evidence-based practice and scientific-based exercise prescription.[9]

Outcome measures have been previously mentioned in two systematic reviews of exercise randomized controlled trials (RCTs) by Klepper and colleagues [2] and Iversen and colleagues.[6] The purpose of the 2019 systematic review was to provide evidence for the safety and efficacy of structured exercise, whereas the 2022 systematic review aimed to provide more detail on the JIA cohort's PA recommendations. Even though these systematic reviews report on the OMs used, they do not delve into the psychometric properties and practicality of these OMs. Psychometric properties refer to whether the RCTs report validity and reliability of the OMs used, as defined by the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) guidelines. Practicality, which includes age-appropriateness, refers to whether the best OM was selected based on the cohort being assessed. Appropriate selection of OMs would contribute to evidence-based exercise prescription for the various physical health signs and symptoms experienced by children diagnosed with JIA.

Children diagnosed with JIA have reduced participation in physical activity secondary to reduced functional capacity and musculoskeletal pain and stiffness. Certain medications may also result in side effects such as weight gain, muscle atrophy, reduced bone density, immunosuppressive effects, and toxicities.[12] Therefore, exercise prescription is beneficial in JIA, but a thorough assessment of the child's physical health status needs to be conducted before exercise prescription can occur. Hence, this systematic review aims to assess the OMs implemented in RCTs and to report on whether they considered the OMs psychometric properties. Furthermore, the identified OMs in the RCTs will then be discussed regarding existing psychometric properties and their age appropriateness.

Material & Methods

Protocol and Registration

The Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) were used to conduct this review.[13] One reviewer (SZ) independently screened studies for eligibility and applied the same search strategy across all databases. All results were gathered, duplicates were removed, and the title and abstracts were examined for inclusion/exclusion criteria. All articles not excluded were combined to conduct a full-text analysis for further inclusion/exclusion, quality assessments following the Joana Briggs Institute (JBI), and data extraction. Additionally, previous exercise systematic reviews within the JIA cohort were reviewed for additional randomized controlled trials to be identified and undergo full-text analysis. Uncertainties from the first reviewer (SZ) regarding inclusion/exclusion and quality assessment were resolved by a second (KD) and third reviewer (KW). The review was registered with PROSPERO (CRD42022336345) on May 30, 2022, before data search and extraction, and all effort was taken to avoid subjective bias.

Data Sources and Search Strategy

Following the preliminary search, the databases PubMed, EBSCOhost, Web of Science, and Ovid were searched from the review commencement with the leading search string of 'juvenile idiopathic arthritis OR childhood arthritis AND exercise* OR physical activity.' The search was initially widened using Boolean operators and wildcards (* and ?) and adjusted according to each engine's specifications. Supplementary Table 1 includes each search string for each specific search engine. Where possible, each keyword was searched under the condition of '[Title/Abstract]' and 'juvenile arthritis' as the main medical subject heading (MeSH term). The search has been updated twice, first on November 20th, 2023, and again on April 11th, 2025.

Inclusion and Exclusion Criteria

Studies included 1) examined children with a definite JIA diagnosis according to the ILAR, 2) both sexes, and 3) in randomized controlled trials. Studies were excluded if they were 1) animal studies, 2) foreign languages, 3) conference papers, 4) grey literature, 5) review articles, 6) non-exercise related, 7) qualitative studies, and 8) psychometric studies.

Data Extraction

Included studies underwent data extraction, specifically sample description specific to the inclusion and exclusion criteria (number of participants per group, age, sex, JIA subtype, and control and interventions) and the OMs implemented (primary and secondary). Additionally, as part of the description of the articles, data related to the JIA subtype, control intervention, and exercise intervention were also collected. Data extraction was expanded to include the type of OMs implemented, the type of health and performance domains assessed by studies, a summary of the outcome measure protocol, and whether the article reported external validity and reliability of the implemented outcome measure. The psychometric data extracted specifically refers to whether the included study stated, either narratively or with an in-text reference, that their OMs used had validity and reliability in a pediatric or JIA population.

Quality and Risk of Bias Assessment

Once consensus through discussion was reached on full-text inclusion, the methodological quality of the studies was appraised using the JBI Tool for RCTs (https://joannabriggs.org/critical-appraisal-tools) and then summarized (Supplementary Table 3). Methodological questions were answered based on whether the article conducted a specific aspect of an RCT ("Yes" or "No"), whether it was unclear ("Unclear"), or whether the question did not apply to the specific study ("Not applicable"). All studies included were evaluated for methodological quality by SZ, KD, and KW to ensure consensus was reached through discussion on the quality of each study. Every study that met the inclusion criteria, independent of their quality, was included. A *priori-determined* criteria of ≤ 40% was set as poor, 41 - 59% as average, 60 – 79% as good, and ≥ 80% as excellent quality score, based on standard lab practice.

Results

Study Selection

Of the four databases searched, 822 articles were identified, of which 387 were duplicates, leaving 435 articles to be screened. Through reading titles and abstracts, a further 409 articles were excluded due to not meeting inclusion criteria. Furthermore, one article could not be retrieved due to no response from the corresponding authors; therefore excluded. The remaining 26 articles were assessed according to the full text's inclusion and exclusion criteria. A further nine articles were excluded due to not being RCTs, not applying an exercise-based intervention, being qualitative, and the exercise intervention not targeting JIA (Figure 1). In addition to the final 16 included articles, four were identified from two previously published systematic reviews within the JIA cohort.[2,6] Hence, 20 full-text articles were assessed and included.[14–32]

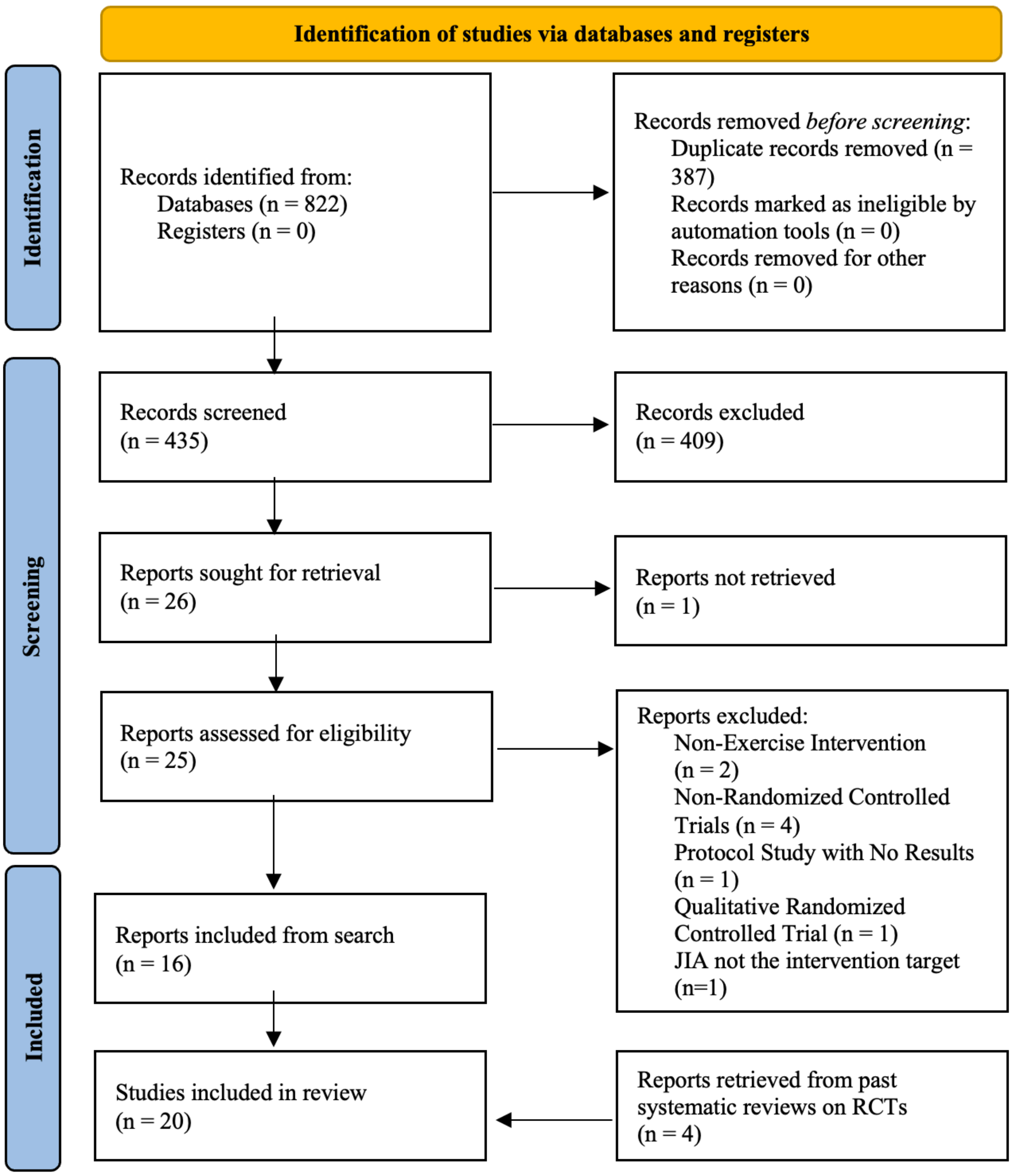

Figure 1: The PRISMA Flow Chart for article selection.

Characteristics of Studies Included

Supplementary Table 2 summarizes the essential study characteristics concerning the specific RCT design (if available), location, relevant inclusion criteria, the population-specific to the JIA subtype, and the experimental and control interventions. Studies were conducted between 2003 and 2024, specifically three in the 2000s,[23,29,31] 10 in the 2010s,[14,16,17,20,24–27,30,32] and seven in the 2020s.[15,18,19,21,22,28,33] Diagnosis, according to ILAR, was the basis of most inclusion criteria across the 20 articles, with only 25% not specifying it.[23,25,26,29,30] The majority of the RCTs had muscle-strengthening interventions (35%),[15,19,22,26–28,30,32] followed by water-based interventions (30%),[17,20,23,25,31,33] and Pilates (15%).[15,18,24] Only one article focused on aerobic,[29] balance,[16] or task-orientated interventions,[14] respectively (5%). Most publications originate out of the Middle East (50%),[14–17,19–22,32,33] followed by Europe (25%),[23,26–28,31] South America (15%),[18,24,25] and North America (10%).[29,30]

Participants

Eight hundred sixty-two participants were included in the 20 studies: 460 in the control group and 493 in the intervention/experimental group. Two articles (11%) did not specify an age of inclusion or exclusion,[20,31] but the 75 remaining articles included participants ranging from five to 21 years of age. In the control group, 163 participants were male and 275 female, and in the exercise group 183 were male and 292 were female. Two articles did not specify the sex of 20 participants in the control group and 18 in the exercise group.[18,20] Out of the 20 articles, seven only included JIA participants diagnosed with polyarticular JIA,[15,15,19,20,22,30,33] whereas the other 13 had different JIA subtypes. Only three articles reported a significant difference between one or two of the descriptive baseline data. Specifically, disease duration was reported to be significantly different in one article,[16] sex in one article,[25] and height and weight in another.[26]

Quality Appraisal

The quality of the 20 exercise RCTs is shown in Supplementary Table 3. The mean score of the exercise RCTs, based on the JBI Critical Appraisal Tool for RCTs, was 8/13 (64%), ranging from five to ten. All articles were identified as not having reported OMs reliably (Question 11, 0%). According to the JBI Critical Appraisal Tool for RCTs, measuring an outcome reliably requires a statement or in-text reference of the OMs psychometric properties, whether assessors were trained, and the number of raters or trials done. Some articles partially adhered to these aspects, but none fully complied with all OMs used. However, all articles measured OMs the same way between treatment groups (Question 10, 100%). Based on the *priori-determined* criteria, one article had poor quality, five average, 14 good, and zero excellent.

Findings of Outcome Measures Implemented

A total of 51 OMs were implemented across the 20 RCT studies, of which two were clinician-reported OMs (CROMs) (4%), 19 were P/PROMs (37%), and 30 were PPOMs (59%), as seen in Supplementary Table 4. The OMs assessed various health and performance domains: disease activity, functional ability, pain, quality of life, fatigue, cardiovascular fitness, range of motion, muscle strength, balance, and anaerobic power. The two articles that used the domain of disease activity each used a different CROM, with one (5%) using the cJADAS[18] and the other (5%) the ACRPedi.[23]

Functional ability, pain, quality of life, and fatigue were assessed using P/PROMs. The most assessed domain was pain, with 13 articles (65%) implementing such an outcome measure. The three P/PROMs used to determine pain included the visual analog scale (VAS) in eight articles (40%),[15,17,20,21,23,24,27,32] the numerical rating scale (NRS) in two articles (10%),[14,16] and the Wong-Baker Face Scale in two articles (10%).[18,30] Functional ability was assessed in 13 studies (65%), with the most common method of assessment being the CHAQ implemented in 11 of these 13 studies[14,16,21,23–26,29–32] and one study implementing the CHAQ-28.[15] Quality of life was the third most commonly assessed domain, with 10 (50%) studies implementing P/PROMs. A high level of variety of the types of P/PROMs used to assess QoL was seen across the 20 articles, with a single article (5%) implementing either PedsQL,[15] EQ-5D,[23] CHQ-C87,[26] VAS,[29] and JAQQ.[31] The PedsQL 3.0 was implemented in three (15%) articles,[18,28,32] while the PedsQL 4.0[24,25] and CHQ-PF50[23,31] were each implemented in two articles (10%) across the 20 included studies. Fatigue was one of the least assessed domains, with only two articles assessing it using P/PROMs, namely the Kids Fatigue Severity Scale (5%)[30] and the PedsQL-MFS (5%).[21]

The most commonly assessed performance domains included muscle strength (50%), cardiorespiratory fitness (CRF) (40%), and range of motion (ROM) (35%). Regarding CRF, five (25%) of these protocols used a cycle ergometer,[15,17,23,30,31] one (5%) treadmill walking,[29] one (5%) the Harvard Step Test,[26] and one (5%) a Bruce Protocol.[28] Range of motion was assessed in four (20%) studies using the pEPM-ROM,[17,24,29,31] in two (10%) studies using a universal goniometer,[16,26] and only one (5%) study assessed ROM using the 10-joints Global Range of Motion Scale (GROMS).[25] Muscle strength was assessed using an isometric handheld dynamometer (HHD)[14,16,23] in three articles (15%), isokinetic testing[19–21,30] in four articles (20%), and isotonic HHD[26] in one article (5%). Isometric techniques also included grip strength[14,26] and pinch strength.[14] Respiratory muscle strength assessments were also included in a single (5%) study.[28] The least assessed performance domains were functional capacity, balance, and anaerobic power. Functional capacity was assessed in six articles (30%), all using the 6-minute walk test (6MWT).[19,22,28,31–33] Both articles[16,26] (10%) that implemented balance assessment used different OMs, one[16] using the flamingo and functional reach test, and the other using only the balance/pediatric reach test.[26] Finally, anaerobic power was only assessed in two studies (10%), one using a full Wingate[17] and the other a modified Wingate protocol.[29]

Reporting of psychometric properties (either narratively or using an in-text reference), namely any form of validity or reliability, was conducted poorly in most studies, as seen in Supplementary Table 4. Only one study[32] provided a full validity report for all its OMs. Eleven out of 20 articles partially reported on the validity of their chosen OMs implemented.[14–16,21,23–26,29,33] Seven articles did not report validity for their chosen OMs.[17,18,20,22,27,30,31] Only eight articles[14,16,21,22,24,26,32] reported reliability partially for their chosen OMs, and ten articles provided no reliability report.[15,17,18,20,23,25,27,29–31]

Discussion

The systematic review aimed to identify the current OMs selected in exercise RCTs and whether studies report on the validity and reliability of their selected OMs. Various OMs are currently implemented in RCTs, but rarely do the RCTs recognize whether they are using valid and reliable OMs. Furthermore, the issue of the report on psychometric properties further creates the question of whether implemented OMs are biopsychosocially appropriate for the JIA cohort.

Clinician Reported Outcome Measures

A discrepancy exists in implementing OMs across RCTs in the JIA cohort, as P/PROMs and PPOMs are more widely implemented than CROMs. Exercise interventions attempt to impact performance domains such as muscle strength, range of motion, and aerobic capacity, and not necessarily clinical domains of disease severity, joint damage, or disease systemic disease manifestations. Consequently, exercise RCTs focused more on PPOMs than CROMs. However, disease activity scores can be crucial determinants of treatment efficacy.[34] Hence, studies should focus more on including CROMs in their outcomes. Epps and colleagues[23] and Calik and colleagues[18] were the only RCTs to implement OMs related to disease activity, the ACRPedi and cJADAS, respectively.

Improvement in physical domains may lead to improvement in clinical domains, as demonstrated by Calik and colleagues.[35] The RCT assessed the impact of Pilates and found participants improved their cJADAS scores and motor ability as measured by the BOT-2SF.[18] Yet, the cause-and-effect relationship between exercise and disease activity still needs to be assessed by implementing applicable CROMs in RCTs. Such implementation will allow researchers to establish whether clinical domains improve as physical domains improve.

Patient- and Parent-Reported Outcome Measures

Including P/PROMs as OMs allows subjective information to be provided as an informal response. However, responses become objective when a questionnaire is developed and determined to be valid and reliable.[36] Furthermore, when collecting information on symptoms, the only accurate source is the patient's perception. Hence, the participant must be asked directly through valid and reliable methods for an objective measure.[36] Understanding the validity of a P/PROM is crucial, particularly concerning age appropriateness, encompassing age-related biopsychosocial development,[37] comprehensiveness of health concepts, and cognitive abilities.[38] Across the JIA cohort, different P/PROMs were implemented to assess pain, physical functioning, and QoL. Inevitably, the P/PROMs used to assess these OMs need to be scrutinized for psychometric appropriateness and validity.

According to Bele and colleagues,[38] participants eight years and older can reliably report their health status despite the challenges faced in developing P/PROMs relating to the age-based biopsychosocial development of children.[37] Pain was assessed using VAS, NRS, and FACES across 12 included exercise RCTs in children and adolescents aged four to 20. Consequently, it is important to understand psychometrically for which ages these rating scales are appropriate when asking children and adolescents to grade chronic pain. The VAS as a measure of pain has a weak recommendation when used in children aged three to six.[39] A weak recommendation relates to insufficient data to measure properties of reliability, content validity, hypothesis testing, cross-cultural validity, criterion validity, and responsiveness, as laid out by the COSMIN guidelines. Furthermore, there is a weak recommendation for using VAS with children aged seven to 18, meaning at least one of the properties assessed is of fair quality. Similarly, depending on their numerical competency, the NRS has a weak recommendation for use in children eight to 18. However, recommendations for using the NRS in children six to eight are inconclusive, meaning there are not enough studies assessing it psychometrically or studies conducted by the same investigators. Hence, difficulty in supporting the selection of the NRS as an assessment tool due to the inconsistency in its psychometric properties. Lastly, the FACES also has inconclusive recommendations for use in children aged eight to 18.[39] Thus, six articles implemented pain rating scales inappropriately related to the age-appropriateness of P/PROMs to measure the outcome of pain.[14,16–18,23,32]

The inclusion of the CHAQ as an outcome measure for physical functioning also relates to subjectivity and age-appropriateness, as previously discussed. The CHAQ can be completed by either the patient when eight years or older or via proxy when younger than eight years old, with age cut-offs appropriate according to Bele and colleagues'[38] proposition regarding instrument validity. Proxy-reported OMs ask caregivers to make inferences about the child's experiences.[37] Hence, it no longer becomes the patient's perception. Such technicality creates the risk of a P/PROMs becoming a subjective assessment rather than an objective assessment, as self-reported and proxy-reported measures have a delicate relationship.[38]

Across the 11 RCTs that included CHAQ, seven[14,16,24,26,29,30,32] did not report whether it was self- or proxy-administered, and it cannot be assumed that they followed the age cut-off. Out of the other four articles, two[21,31] conducted CHAQs that were proxy-reported, one[25] conducted a self-reported CHAQ, and one[23] conducted both a child- and proxy-reported CHAQ. When administering the CHAQ appropriately, it has good test-retest reliability (ICC of 0.82), good to excellent internal consistency (Cronbach's alpha ranging from 0.88 to 0.96), acceptable interrater reliability ($r = 0.54$-0.84, $p < 0.05$), and confirmed face validity.[40]

Self-perceived measures such as QoL should be self-reported by children if the administered questionnaire permits it concerning age-appropriateness. Administration and validity should be well understood when implementing an assessment instrument for QoL. Assessing the administration and age appropriateness of the instruments implemented in the JIA cohorts brings into question the continuity of instruments implemented. Eight instruments were used across nine studies that assessed QoL as an outcome. Two of these eight instruments were developed explicitly for the JIA cohort: the Juvenile Arthritis Quality of Life Questionnaire (JAQQ)[31] and the Pediatric Quality of Life Inventory 3.0 (PedsQL 3.0) Rheumatology

Module.[18,28,32] Both instruments have been validated for the cohort, with possible self-reported or proxy-reported administration. However, age-appropriateness concerns the length of the JAQQ, as it contains 74 items across five domains, making it time-consuming[41] and challenging regarding cognitive abilities concerning concentration in children. Hence, the length and time of the JAQQ give rise to practical issues when implementing such a measure within a clinical setting. The PedsQL 3.0 has sufficient validity and excellent reliability, whether self-reported or proxy-reported, making it a strong objective and rheumatic-specific instrument to assess QoL.[42]

The remaining six instruments contained five pediatric-specific QoL questionnaires and one general assessment of QoL. The general assessment refers to the VAS used to grade QoL within the CHAQ Discomfort Index,[25] which raises the question of age-appropriateness about understanding the numerical grading scale concerning symptoms and emotions. Specifically, it has been found that children find it easier to report on observable behaviors rather than emotions.[43] Yet, QoL questionnaires ask children to consider their feelings and how their illness impacts their lives.[44] Similarly, QoL questionnaires specific to the pediatric cohort also ask questions about emotional functioning, specifically in the variations of the PedsQL,[15,21,24,25] CHQ-PF50,[23,31] EQ-5D,[23] and CHQ-C87.[26]

Physical Performance Outcome Measures

Cardiorespiratory fitness (CRF) was assessed using cycling,[15,17,23,30,31] stepping,[26] and walking assessments[22,29,31,32] across eight included exercise RCTs. However, continuity of protocol use could be improved, with each RCT using a different assessment method or the same but different protocols. Five RCTs conducted cycle ergometer assessments, but each implemented protocol differed from one investigation to the next. Only two articles mentioned established protocols they followed, namely the McMaster Incremental Protocol[15] and an adapted version of the Giannini and Protas Protocol.[23] No protocols were mentioned in three articles. However, two did indicate how they assessed the revolutions per minute, initial loading, and timing of incremental loading.[17,31] The last article only stated the specific CRF outcome measure, VO_2 peak, but did not mention a protocol.[30]

The McMaster Incremental Protocol is a recommended fitness test for young individuals.[45] It is considered appropriate for implementation in children aged 10 to 14, as done by Azab and colleagues.[15] Specifically, the protocol considers the child's height to determine initial loading, incremental loading, and duration between loads.[46] Similarly, the Giannini and Protas protocol[47] was developed from the James All-Out Progressive Continuous Cycling Test, which has also been deemed appropriate for fitness testing in children.[46] Here, a child's body surface area determines initial, incremental loading, and duration between loads.[46,47] However, Epps and colleagues implemented this assessment in children as young as four.[23] Regardless of children starting at a comfortable rate of pedaling, the starting wattage and incremental increases are equal to that of adult protocols, such as the YMCA,[48] therefore too high. This raises the question of the age-appropriateness of the protocol used by Epps and colleagues for the age of their participants. However, provision was made for active joints and reduced range of motion to ensure a 15-degree flexion at the lower part of the cycling,[23]

Similarly, the protocol used by Bayraktar and colleagues raises questions of age-appropriateness, as this protocol is implemented for participants as young as eight at an initial load of 50 wattage with an incremental increase of 25 wattage every minute. The increments are the same as that used in an adult protocol such as the YMCA, and the starting load is double that.[48] However, Takken and colleagues[31] implemented more appropriate initial loads of zero Watts and a lower incremental load increase and duration of 20 wattage every three minutes.

Beyond cycle fitness tests, other fitness assessments included two treadmill tests,[28,29] the Harvard Step Test,[26] and the 6MWT in six different studies.[19,22,28,31–33] One treadmill protocol[29] and the Harvard Step Test has not been validated in the pediatric population. The Bruce protocol has been validated in a pediatric population,[49] but the large incremental increases in workload need to be reconsidered for a clinical population. Such large incremental increases, especially in gradient, in cohorts with joint pathology, may exacerbate joint pain and lead to premature termination of the test. The latter may lead to inaccurate measurements of CRF due to orthopedic limitations.

Recent efforts have been made to develop the 6MWT within the JIA cohort. A low-to-moderate validity of the 6MWT has been reported in correlation with VO_2peak, as the 6MWT may be more indicative of joint status than aerobic capacity.[50,51] More recently, the reproducibility of the 6MWT in the JIA population for children aged seven to 17 has been explored. Pritchard and colleagues[51] found that the 6MWT displays good to excellent reliability (ICC = 0.86 with 95% confidence interval) in the JIA population, with a smallest detectable difference of 65.1 meters. Not only has 6MWT's validity and reliability been established within the JIA population, but reference values with a predictive model have been established. These psychometric properties of the 6MWT in the JIA population have been established using the American Thoracic Society (ATS) guidelines of the 6MWT procedure, with excellent reliability demonstrated (ICC = 0.86).[50–52]

When assessing the implementation of the 6MWT in the four included exercise RCTs, only two[22,33] maintained the ATS guidelines, as they used a walking distance of 30 meters. However, Mian and colleagues[53] used a 25-meter walkway with the ATS guidelines to determine reference values in the JIA cohort. Elnagger and colleagues[22,33] also appropriately implemented the 6MWT regarding its validation and reliability associated with age, as the test was conducted with participants aged 10 to 18 years. Lastly, the investigations by Takken and colleagues[31] and Tarakci and colleagues[32] used eight-meter walkways, which do not follow the ATS guidelines, and implemented the assessment in children younger than seven.

Beyond CRF assessments, the implementation of anaerobic power tests and balance assessments must also be considered within the JIA cohort for age-appropriateness and lack of implementation. A Wingate[17] and a modified Wingate[29] have been implemented in the JIA cohort to assess anaerobic power. However, only the modified Wingate has been evaluated for reliability, with power at 10 seconds in wattage having an ICC of 0.92 and power at 30 seconds in wattage having an ICC of 0.94.[54] Furthermore, reliability does not imply validity, which encompasses the age-appropriateness of an assessment. Hence, implementing both the Wingate and modified Wingate needs to be reconsidered until further psychometric testing is done.

Concerning the use of balance tests in the JIA cohort, only two articles[16,26] included such measurements. Inflammation and joint destruction may also alter neuromuscular function in a child diagnosed with JIA. Neuromuscular function depends on sensory input from proprioceptors, vision, and the vestibular system to initiate neuromuscular responses.[55] Proprioception depends on mechanoreceptors in the joint capsule, ligaments, tendons, and skin, providing input for arthokinetic and muscular reflexes to maintain balance and postural control. Pro-inflammatory markers within a joint may result in the destruction or alteration of mechanoreceptors. Hence, children with JIA may experience balance perturbations from proprioceptive impairment and deficits.[55] Consequently, a balanced evaluation concerning a child's physical health status is crucial.

Limitations, Clinical Implications, and Future Directions

Limitations and Strengths

There is a recognition of limitations regarding the number of reviewers who conducted the search and determined the inclusion and exclusion of articles, as this increases the risk of bias. However, strengths include that discrepancies of inclusion were resolved by a second reviewer, with three reviewers conducting quality appraisals on the included studies. Lastly, the systematic review was registered on PROSPERO and followed the PRISMA guidelines.

Clinical Implications and Future Directions

Clinicians need to focus more on implementing OMs that have been validated and reliable within a pediatric population. Outcome measures should also not only focus on performance, but also on clinical aspects such as disease activity. Age-appropriateness and practicality must be considered in OM selection, especially regarding the length and time of P/PROMs and whether children can appropriately meet the physical capacities required of the specific PPOMs selected. Therefore, reporting of psychometric properties for implemented OMs in RCTs needs to be improved. Inclusion of a pediatric scientist in the research team may also be beneficial to avoid the use of adult-based protocols.

Consequently, future research needs to focus on whether a cause-and-effect relationship exists between physical performance and clinical outcomes while using valid and reliable OMs. A standardized, holistic group of OMs can be established through future research to assess the physical health status of children with JIA. Such research would allow for better comparison across interventions in research, and also assessment of physical domains not regularly included in research, such as proprioception and postural control.

Conclusion

A wide variety of OMs have been implemented within the research of the JIA cohort. Hence, it is difficult to compare across different interventions, and creates a practical difficulty in selecting OMs in clinical practice. Furthermore, more focus should be placed on how exercise improves clinical outcomes such as disease activity and severity. Hence, CROMs need to be explored more in conjunction with PPOMs. Additionally, it should be ensured that the correct protocol of P/PROMs are implemented within the JIA cohort to maintain validity, age-appropriateness, and practicality of the implemented P/PROM. Lastly, PPOMs need to be more disease-specific and directed at children's health needs, such as improving their joint functioning and systemic health. Implementing adult-like PPOMs, such as the cycle ergometer assessments and anaerobic power assessments, should be reconsidered for the JIA cohort with active disease.

References

1. Martini A, Ravelli A, Avcin T, et al. Toward new classification criteria for juvenile idiopathic arthritis: first steps, Pediatric Rheumatology International Trials Organization International consensus. *J Rheumatol*. 2019;46(2):190-197. doi:10.3899/jrheum.180168

2. Klepper S, Mano Khong TT, Klotz R, Gregorek AO, Chan YC, Sawade S. Effects of structured exercise training in children and adolescents with juvenile idiopathic arthritis. *Pediatr Phys Ther*. 2019;31(1):3-21. doi:10.1097/PEP.0000000000000555

3. Pritchard L, Verschuren O, Roy M, Kaup C, Rumsey DG. Reproducibility of the Six-Minute Walk Test in Children and Youth with Juvenile Idiopathic Arthritis. *Arthritis Care & Research*. 2022;74(4):686-690. doi:10.1002/acr.24492

4. Al-Mayouf SM, Al Mutairi M, Bouayed K, et al. Epidemiology and demographics of juvenile idiopathic arthritis in Africa and Middle East. *Pediatr Rheumatol*. 2021;19(1):166. doi:10.1186/s12969-021-00650-x

5. Ringold S, Angeles-Han ST, Beukelman T, et al. 2019 American College of Rheumatology/Arthritis Foundation Guideline for the Treatment of Juvenile Idiopathic Arthritis: Therapeutic Approaches for Non-Systemic Polyarthritis, Sacroiliitis, and Enthesitis. *Arthritis Care Res*. 2019;71(6):717-734. doi:10.1002/acr.23870

6. Iversen MD, Andre M, von Heideken J. Physical activity interventions in children with juvenile idiopathic arthritis: a systematic review of randomized controlled trials. *Pediatric Health Med Ther*. 2022;14(13):115-143. doi:10.2147/PHMT.S282611

7. Kuntze G, Nesbitt C, Whittaker JL, et al. Exercise therapy in juvenile idiopathic arthritis: a systematic review and meta-analysis. *Arch Phys Med Rehabil*. 2018;99(1):178-193. doi:10.1016/j.apmr.2017.05.030

8. Wedge FM, Braswell-Christy J, Brown CJ, Foley KT, Graham C, Shaw S. Factors influencing the use of outcome measures in physical therapy practice. *Physiother Theory Pract*. 2012;28(2):119-133. doi:10.3109/09593985.2011.578706

9. Al-Muqiren TN, Al-Eisa ES, Alghadir AH, Anwer S. Implementation and use of standardized outcome measures by physical therapists in Saudi Arabia: barriers, facilitators and perceptions. *BMC Health Serv Res*. 2017;17(1):748. doi:10.1186/s12913-017-2693-2

10. Fetters L, Tilson J. *Evidence Based Physical Therapy.* 2e ed. FA Davies Company; 2019.

11. Griffiths A, Toovey R, Morgan PE, Spittle AJ. Psychometric properties of gross motor assessment tools for children: a systematic review. *BMJ Open*. 2018;8(10):e021734. doi:10.1136/bmjopen-2018-021734

12. Acer Kasman S, Karaahmet ÖZ, Bal Hastürk A, et al. The importance of lower extremity involvement type on exercise performances, function, and quality of life in peripheral predominant forms of juvenile idiopathic arthritis. *Adv Rheumatol*. 2021;61(1):39. doi:10.1186/s42358-021-00195-3

13. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine*. 2009;151(4):264-269.

14. Arman N, Tarakci E, Tarakci D, Kasapcopur O. Effects of video games–based task-oriented activity training (Xbox 360 Kinect) on activity performance and participation in patients with juvenile idiopathic arthritis: a randomized clinical trial. *Am J Phys Med Rehabil*. 2019;98(3):174-181. doi:10.1097/PHM.0000000000001001

15. Azab AR, Kamel FH, Basha MA, et al. Impact of clinical Pilates exercise on pain, cardiorespiratory fitness, functional ability, and quality of life in children with polyarticular juvenile idiopathic arthritis. *IJERPH*. 2022;19(13):7793. doi:10.3390/ijerph19137793

16. Baydogan SN, Tarakci E, Kasapcopur O. Effect of strengthening versus balance-proprioceptive exercises on lower extremity function in patients with juvenile idiopathic arthritis: a randomized, single-blind clinical trial. *Am J Phys Med Rehabil* . 2015;94(6):417-428. doi:10.1097/PHM.0000000000000279

17. Bayraktar D, Savci S, Altug-Gucenmez O, et al. The effects of 8-week water-running program on exercise capacity in children with juvenile idiopathic arthritis: a controlled trial. *Rheumatol Int*. 2019;39(1):59-65. doi:10.1007/s00296-018-4209-8

18. Calık BB, Gur Kabul E, Korkmaz C, Tekin ZE, Yener GO, Yuksel S. The efficacy of clinical Pilates exercises in children and adolescents with juvenile idiopathic arthritis: a pilot study. *Rev Colomb de Reumatol*. 2020;27(4):269-277. doi:10.1016/j.rcreu.2020.06.015

19. Elnaggar RK, Azab AR, Alrawaili SM, et al. Efficacy of accommodating variable-resistance training on muscle architecture, peak torque, and functional performance in patients with juvenile idiopathic arthritis: A randomized controlled trial. *Heliyon*. 2024;10(6):e27693. doi:10.1016/j.heliyon.2024.e27693

20. Elnaggar RK, Elshafey MA. Effects of combined resistive underwater exercises and interferential current therapy in patients with juvenile idiopathic arthritis: a randomized controlled trial. *Am J Phys Med Rehabil*. 2016;95(2):96-102. doi:10.1097/PHM.0000000000000347

21. Elnaggar RK, Mahmoud WS, Abdelbasset WK, Alqahtani BA, Alrawaili SM, Elfakharany MS. Low-energy laser therapy application on knee joints as an auxiliary treatment in patients with polyarticular juvenile idiopathic arthritis: a dual-arm randomized clinical trial. *Lasers Med Sci*. 2022;37(3):1737-1746. doi:10.1007/s10103-021-03427-6

22. Elnaggar RK, Mahmoud WS, Moawd SA, Azab AR. Impact of core stability exercises on bone mineralization and functional capacity in children with polyarticular juvenile idiopathic arthritis: a randomized clinical trial. *Clin Rheumatol*. 2021;40(1):245-253. doi:10.1007/s10067-020-05219-9

23. Epps H, Ginnelly L, Utley M, et al. Is hydrotherapy cost-effective? *Health Technology Assessment*. 2005;9(39).

24. Mendonça TM, Terreri MT, Silva CH, et al. Effects of Pilates exercises on health-related quality of life in individuals with juvenile idiopathic arthritis. *Archives of Physical Medicine and Rehabilitation*. 2013;94(11):2093-2102. doi:10.1016/j.apmr.2013.05.026

25. Pérez Ramírez N, Nahuelhual Cares P, San Martín Peñailillo P. Efectividad de la terapia Watsu en pacientes con artritis idiopática juvenil. Un ensayo clínico controlado paralelo, aleatorio y simple ciego. *Rev Chil Pediatr*. 2019;90(3):282. doi:10.32641/rchped.v90i3.886

26. Sandstedt E, Fasth A, Eek MN, Beckung E. Muscle strength, physical fitness and well-being in children and adolescents with juvenile idiopathic arthritis and the effect of an exercise programme: a randomized controlled trial. *Pediatr Rheumatol Online J*. 2013;11(1):7. doi:10.1186/1546-0096-11-7.

27. Sandstedt E, Fasth A, Fors H, Beckung E. Bone health in children and adolescents with juvenile idiopathic arthritis and the influence of short-term physical exercise. *Pediatric Physical Therapy*. 2012;24(2):155-161. doi:10.1097/PEP.0b013e31824cce6e

28. Sarac DC, Bayraktar D, Ozer Kaya D, Altug Gucenmez O, Oskay D. The effects of inspiratory muscle training on cardiorespiratory functions in juvenile idiopathic arthritis: A randomized controlled trial. *Pediatric Pulmonology*. 2024;59(3):562-573. doi:10.1002/ppul.26783

29. Singh-Grewal D, Schneiderman-Walker J, Wright V, et al. The effects of vigorous exercise training on physical function in children with arthritis: A randomized, controlled, single-blinded trial. *Arthritis Rheum*. 2007;57(7):1202-1210. doi:10.1002/art.23008

30. Sule S, Fontaine K. Slow speed resistance exercise training in children with polyarticular juvenile idiopathic arthritis. *OARRR*. 2019;Volume 11:121-126. doi:10.2147/OARRR.S199855

31. Takken T. Aquatic fitness training for children with juvenile idiopathic arthritis. *Rheumatology*. 2003;42(11):1408-1414. doi:10.1093/rheumatology/keg386

32. Tarakci E, Yeldan I, Baydogan S, Olgar S, Kasapcopur O. Efficacy of a land-based home exercise programme for patients with juvenile idiopathic arthritis: A randomized, controlled, single-blind study. *J Rehabil Med*. 2012;44(11):962-967. doi:10.2340/16501977-1051

33. Elnaggar RK, Elfakharany MS. Aqua-plyometric exercises-induced changes in muscle strength, bone mineral properties, and physical fitness in patients with juvenile idiopathic arthritis: a 12-week, randomized controlled trial. *Pediatr Exerc Sci*. 2023;35(4):198-205. doi:10.1123/pes.2022-0044

34. Ringold S, Consolaro A, Ardoin SP. Outcome measures in pediatric rheumatic disease. *Rheum Dis Clin North Am*. 2021;47(4):655-668. doi:10.1016/j.rdc.2021.07.013

35. Calık BB, Gur Kabul E, Korkmaz C, Tekin ZE, Yener GO, Yuksel S. The efficacy of clinical Pilates exercises in children and adolescents with juvenile idiopathic arthritis: A pilot study. *Revista Colombiana de Reumatología*. 2020;27(4):269-277. doi:10.1016/j.rcreu.2020.06.015

36. Hamilton DF, Giesinger JM, Giesinger K. It is merely subjective opinion that patient-reported outcome measures are not objective tools. *Bone & Joint Research*. 2017;6(12):665-666. doi:10.1302/2046-3758.612.BJR-2017-0347

37. Kwon J, Freijser L, Huynh E, et al. Systematic Review of Conceptual, Age, Measurement and Valuation Considerations for Generic Multidimensional Childhood Patient-Reported Outcome Measures. *PharmacoEconomics*. 2022;40(4):379-431. doi:10.1007/s40273-021-01128-0

38. Bele S, Chugh A, Mohamed B, Teela L, Haverman L, Santana MJ. Patient-Reported Outcome Measures in Routine Pediatric Clinical Care: A Systematic Review. *Front Pediatr*. 2020;8:364. doi:10.3389/fped.2020.00364

39. Birnie KA, Hundert AS, Lalloo C, Nguyen C, Stinson JN. Recommendations for selection of self-report pain intensity measures in children and adolescents: a systematic review and quality assessment of measurement properties. *Pain*. 2019;160(1):5-18. doi:10.1097/j.pain.0000000000001377

40. Greer AE, Iversen MD. Measures of pediatric function and physical activity in arthritis. *Arthritis Care Res*. 2020;72(Suppl 10):499-521. doi:10.1002/acr.24239

41. Weiss PF, Colbert RA, Xiao R, et al. Development and retrospective validation of the Juvenile Spondyloarthritis Disease Activity Index: Juvenile SpA Disease Activity Index. *Arthritis Care Res*. 2014;66(12):1775-1782. doi:10.1002/acr.22411

42. Młyńczyk J, Abramowicz P, Stawicki MK, Konstantynowicz J. Non-disease specific patient-reported outcome measures of health-related quality of life in juvenile idiopathic arthritis: a systematic review of current research and practice. *Rheumatol Int*. 2022;42:191-203. doi:https://doi.org/10.1007/s00296-021-05077-x

43. Conijn JM, Smits N, Hartman EE. Determining at What Age Children Provide Sound Self-Reports: An Illustration of the Validity-Index Approach. *Assessment*. 2020;27(7):1604-1618. doi:10.1177/1073191119832655

44. Kreimeier S, Greiner W. EQ-5D-Y as a Health-Related Quality of Life Instrument for Children and Adolescents: The Instrument's Characteristics, Development, Current Use, and Challenges of Developing Its Value Set. *Value in Health*. 2019;22(1):31-37. doi:10.1016/j.jval.2018.11.001

45. Brehm MA, Balemans ACJ, Becher JG, Dallmeijer AJ. Reliability of a Progressive Maximal Cycle Ergometer Test to Assess Peak Oxygen Uptake in Children With Mild to Moderate Cerebral Palsy. *Physical Therapy*. 2014;94(1):121-128. doi:10.2522/ptj.20130197

46. Bar-Or O, Rowland T W. *Pediatric Exercise Medicine: From Physiologic Principles to Health Care Application*. Human Kinetics; 2004.

47. Jasso Giannini M, Protas EJ. Exercise Response in Children with and without Juvenile Rheumatoid Arthritis: A Case-Comparison Study. *Physical Therapy*. 1992;72(5):365-372. doi:10.1093/ptj/72.5.365

48. Liguori G, American College of Sports Medicine ACSM. *ACSM's Guidelines for Exercise Testing and Prescription*. 11th ed. Wolters Kluwer; 2021.

49. Gumming GR, Everatt D, Hastman L. Bruce treadmill test in children: Normal values in a clinic population. *The American Journal of Cardiology*. 1978;41(1):69-75. doi:10.1016/0002-9149(78)90134-0

50. Lelieveld OTHM, Takken T, van der Net J, van Weert E. Validity of the 6-minute walking test in juvenile idiopathic arthritis. *Arthritis Rheum*. 2005;53(2):304-307. doi:10.1002/art.21086

51. Pritchard L, Verschuren O, Roy M, Kaup C, Rumsey DG. Reproducibility of the six minute walk test in children and youth with juvenile idiopathic arthritis. *Arthritis Care Res*. 2022;74(4):686-690. doi:10.1002/acr.24492

52. Mian Q, Rumsey DG, Verschuren O, et al. Reference Values for the Six Minute Walk Test in Children with Juvenile Idiopathic Arthritis. *Physical & Occupational Therapy In Pediatrics*. Published online June 24, 2021:1-11. doi:10.1080/01942638.2021.1934239

53. Mian Q, Rumsey DG, Verschuren O, et al. Reference values for the six minute walk test in children with juvenile idiopathic arthritis. *Phys Occup Ther Pediatr*. 2021;42(2):187-197. doi:10.1080/01942638.2021.1934239

54. Stephens S, Singh-Grewal D, Bar-Or O, et al. Reliability of exercise testing and functional activity questionnaires in children with juvenile arthritis. *Arthritis Rheum*. 2007;57(8):1446-1452. doi:10.1002/art.23089

55. Houghton KM, Guzman J. Evaluation of static and dynamic postural balance in children with juvenile idiopathic arthritis. *Pediatr Phys Ther*. 2013;25(2):150-157. doi:10.1097/PEP.0b013e31828a2978