The Six-Minute Walk Test in Chronic Pediatric Conditions: A Systematic Review of Measurement Properties

Bart Bartels, Janke F. de Groot, Caroline B. Terwee

Background. The Six-Minute Walk Test (6MWT) is increasingly being used as a functional outcome measure for chronic pediatric conditions. Knowledge about its measurement properties is needed to determine whether it is an appropriate test to use.

Purpose. The purpose of this study was to systematically review all published clinimetric studies on the 6MWT in chronic pediatric conditions.

Data Sources. The databases MEDLINE, EMBASE, CINAHL, PEDro, and SPORTDiscus were searched up to February 2012.

Study Selection. Studies designed to evaluate measurement properties of the 6MWT in a chronic pediatric condition were included in the systematic review.

Data Extraction. The methodological quality of the included studies and the measurement properties of the 6MWT were examined.

Data Synthesis. A best evidence synthesis was performed on 15 studies, including 9 different chronic pediatric conditions. Limited evidence to strong evidence was found for reliability in various chronic conditions. Strong evidence was found for positive criterion validity of the 6MWT with peak oxygen uptake in some populations, but negative criterion validity was found in other populations. Construct validity remained unclear in most patient groups because of methodological flaws. Little evidence was available for responsiveness and measurement error. Studies showed large variability in test procedures despite existing guidelines for the performance of the 6MWT.

Limitations. Unavailability of a specific checklist to evaluate the methodological quality of clinimetric studies on performance measures was a limitation of the study.

Conclusions. Evidence for measurement properties of the 6MWT varies largely among chronic pediatric conditions. Further research is needed in all patient groups to explore the ability of the 6MWT to measure significant and clinically important changes. Until then, changes measured with the 6MWT should be interpreted with caution. Future studies or consensus regarding modified test procedures in the pediatric population is recommended.

B. Bartels, PPT, MSc, Child Development and Exercise Center, Wilhelmina Children's Hospital, University Medical Center Utrecht, Utrecht, the Netherlands. Address all correspondence to Mr Bartels at: b.bartels-4@umcutrecht.nl.

J.F. de Groot, PT, PhD, Research Group Lifestyle and Health, HU University of Applied Sciences, Utrecht, the Netherlands, and Child Development and Exercise Center, Wilhelmina Children's Hospital, University Medical Center Utrecht.

C.B. Terwee, PhD, Department of Epidemiology and Biostatistics, EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, the Netherlands.

[Bartels B, de Groot JF, Terwee CB. The Six-Minute Walk Test in chronic pediatric conditions: a systematic review of measurement properties. *Phys Ther.* 2013;93:529–541.]

Published Ahead of Print: November 15, 2012
Accepted: November 2, 2012
Submitted: May 14, 2012

The Six-Minute Walk Test (6MWT) is a self-paced walking test generally used to assess functional capacity in people with chronic conditions. The main outcome is the distance that a person can walk in 6 minutes.[1] The 6MWT was originally developed to measure the submaximal level of functional capacity in adult patients with moderate to severe heart or lung diseases and has been extensively used in other patient populations.[2–6] Because the test reflects an exercise level close to that of daily life activities, it is easy to administer, is well tolerated by patients, and is increasingly being used as a functional outcome measure for people with chronic conditions, including pediatric populations. For example, the 6MWT has been utilized as a functional outcome measure in recent intervention studies including children with mucopolysaccharidosis type 1,[7] Duchenne muscular dystrophy,[8] spina bifida,[9] and obesity.[10] Moreover, the 6MWT is increasingly being used and recommended in physical therapist practice.[11] To determine the suitability of the 6MWT as an appropriate functional outcome measure in the pediatric populations, however, knowledge of its measurement properties is needed with respect to the population of interest. The 6MWT's measurement properties include: (1) reproducibility and (2) validity.

Reproducibility refers to the degree of similarity between repeated measurements and reflects both reliability and agreement parameters.[12] Reliability comprises the proportion of the total variance in the measurements that is due to "true" differences among individuals and is expressed as the intraclass correlation coefficient (ICC).[13] Measurement error assesses the intra-individual variability between repeated measurements and often is expressed as the coefficient of variation (CV), the smallest detectable change (SDC) (also referred to as the minimal detectable change [MDC]), and the limits of agreement (LoA).

Validity refers to the degree to which an instrument measures the construct it purports to measure and comprises hypothesis testing (construct validity), criterion validity, and responsiveness. *Hypothesis testing* refers to the degree to which the scores of an instrument are consistent with predefined hypotheses regarding relationships to scores of other instruments (convergent validity) or differences among relevant groups (discriminative validity). *Criterion validity* indicates the degree to which the scores of an instrument are an adequate reflection of a gold standard. *Responsiveness* reflects the ability of an instrument to detect change over time in the construct to be measured and is assessed by testing prespecified hypotheses about the relationship between the change scores of the instrument and changes in other measures.[13]

Next to these statistical measurement properties, the minimal important change (MIC), also referred to as the minimal clinically important difference (MCID), is an important, but often overlooked, property. It refers to the smallest change that is considered relevant by patients and can be assessed both by methods focused on patient perspective and by statistical methods.[14] Knowledge about the MIC of an outcome measure in a specific patient group makes it possible to investigate whether the measurement error of an instrument is small enough to identify relevant changes.

Measurement properties of the 6MWT have been reported for various pediatric populations, but no systematic review has been performed on the methodological quality of these clinimetric studies. Consequently, the level of evidence for the quality of measurement properties and appropriateness of the 6MWT in chronic pediatric conditions is still unknown. Therefore, the aims of this systematic review were to determine the current level of evidence for the measurement properties of the 6MWT in chronic pediatric conditions and to give an overview of available measurement properties in different pediatric populations with a chronic condition. More knowledge on the reproducibility and validity in relation to the MIC of the 6MWT will enable the clinician to determine in which pediatric conditions and for what purpose the 6MWT is appropriate to use in his or her daily physical therapist practice.

Method

Data Sources and Searches

A search was performed up to February 2012 in MEDLINE, EMBASE, CINAHL, PEDro, and SPORTDiscus. In PubMed, a validated search filter for finding studies on measurement properties was applied.[15] The full search strategy is described in eAppendix 1 (available at ptjournal.apta.org).

Study Selection

The selection of the articles was independently performed by 2 reviewers (B.B. and J.F.dG.). The following inclusion criteria were used: (1) the aim of the study was to develop or evaluate measurement properties of the 6MWT, and (2) the 6MWT was evaluated in children with chronic conditions.

Available With This Article at ptjournal.apta.org

- eAppendix 1: Search Filter for PubMed
- eAppendix 2: Methodological Quality of the Studies

Table 1.
Six-Minute Walk Test Procedures Compared With American Thoracic Society (ATS) Guidelines

Study	Location		Instructions				Encouragement	
	Straight Corridor	Length Walking Course (m)	Pretesting Resting Period	Walk As Far As Possible Without Jogging or Running	Turning Around the Cones	Practice Test	Standardized	Phrases in ATS Guidelines
ATS guidelines[1]	Yes	30	Yes	Yes	Yes	Optional	Yes	Yes
Do Santos Alves and Avanzi[40]	Yes	30	No	No	No	No	Yes	Yes
Balfour-Lynn et al[23]	Yes	17	No	No	No	No	No	No
Chong et al[28]	No	25	No	No	No	No	Yes	No
Cunha et al[25]	Yes	28	No	Yes	No	No	No	No
de Groot et al[35]	Yes	20	No	Yes	No	No	Yes	Yes
Elloumi et al[29]	Yes	30	Yes	Yes	No	No	Yes	Yes
Guinhouya[32]	Yes	30	No	Yes	No	No	Yes	Yes
Gulmans et al[22]	Yes	8	No	Yes	No	Yes	Yes	No
Lammers et al[42]	Yes	30–50	No	Yes	No	No	No	No
Lelieveld et al[33]	Yes	8	No	Yes	No	No	Yes	No
Maher et al[26]	Yes	10	Yes	Yes	Yes	No	Yes	Yes
Makni et al[30]	Yes	30	Yes	Yes	Yes	No	Yes	Yes
Mandrusiak et al[24]	Yes	30	Yes	Yes	Yes	Yes	Yes	Yes
Mazzone et al[36]	Yes	30	Yes	Yes	Yes	No	No	No
McDonald et al[38]	Yes	25	Yes	Yes	Yes	No	Yes	No
McDonald et al[37]	Yes	25	Yes	Yes	Yes	No	Yes	No
Moalla et al[39]	Yes	30	Yes	Yes	No	No	Yes	No
Montes et al[41]	Yes	25	No	Yes	Yes	No	Yes	Yes
Morinder et al[31]	Yes	70	No	Yes	No	No	No	No
Paap et al[34]	Yes	8	No	Yes	No	No	Yes	No
Takken et al[43]	Yes	20	No	Yes	No	No	Yes	No
Thompson et al[27]	No	20 × 45	Yes	Yes	No	No	Yes	No

Childhood obesity was considered a chronic condition, given the long-term medical consequences involved.[16] Studies were excluded if adult patients were included in the study sample or if the patient population was not clearly described (eg, "chronically ill patients"). Studies that were not available in full text or were published in a language other than English or Dutch also were excluded.

Data Extraction and Quality Assessment

The extraction and assessment consisted of several steps. First, the testing procedures of the 6MWT were evaluated and compared with conventional guidelines. Second, the methodological quality of the studies was assessed. Third, the quality of the measurement properties of the 6MWT was evaluated. Ultimately, a best evidence synthesis was performed, taking both the methodological quality of the studies and the quality of the measurement properties into account.

Comparison of 6MWT testing procedures (Tab. 1). One reviewer (B.B.) evaluated the 6MWT test procedures used in the included studies and compared them with the American Thoracic Society (ATS) guidelines[17] on the following aspects: location, instructions, and encouragement.

Assessment of the methodological quality of the included studies. Two reviewers (B.B. and J.F.dG.) independently evaluated the methodological quality of the included studies using the recently developed COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist (eAppendix 2, available at ptjournal.apta.org).[13,18] The

Table 2.
Quality Criteria for Measurement Properties[a]

Measurement Property	Positive	Indeterminate	Negative
Reliability	ICC/weighted kappa ≥0.70 or Pearson *r* ≥.80	Neither ICC/weighted kappa nor Pearson *r* determined	ICC/weighted kappa <0.70 or Pearson *r* <.80
Measurement error	MIC > SDC or MIC outside LoA	MIC not determined	MIC ≤ SDC or MIC equals or inside LoA
Hypothesis testing	Correlation with an instrument measuring the same construct ≥0.50 or at least 75% of the results are in accordance with the hypotheses and correlation with related constructs is higher than with unrelated constructs	Solely correlations determined with unrelated constructs	Correlation with an instrument measuring the same construct <0.50 or <75% of the results are in accordance with the hypotheses or correlation with related constructs is lower than with unrelated constructs
Criterion validity	Convincing arguments that gold standard is "gold" and correlation with gold standard ≥0.70	No convincing arguments that gold standard is "gold" or doubtful design or method	Correlation with gold standard <0.70, despite adequate design and method
Responsiveness	Correlation with an instrument measuring the same construct ≥0.50 or at least 75% of the results are in accordance with the hypotheses or AUC ≥0.70 and correlation with related constructs is higher than with unrelated constructs	Solely correlations determined with unrelated constructs	≤75% of the results are in accordance with the hypotheses or AUC <0.70 or correlation with related constructs is lower than with unrelated constructs

[a] ICC=intraclass coefficient, MIC=minimal important change, SDC=smallest detectable change, LoA=limits of agreement, AUC=area under the receiver operating characteristic curve.

COSMIN checklist is a standardized tool for assessing the methodological quality of studies on measurement properties and is designed to evaluate health-related patient-reported outcomes, but also can be used to evaluate measurement properties of other outcome measures.

The COSMIN checklist is a modular tool and contains 12 boxes: 10 boxes can be used to assess the methodological quality of studies on measurement properties, and 2 boxes contain general requirements (refer to the COSMIN checklist[18] for a description of the alphabetic labeling of boxes). Seven boxes on methodological quality were used to assess the quality of the studies regarding reliability, measurement error, content validity, hypothesis testing, cross-cultural validity, criterion validity, and responsiveness. The measurement properties "internal consistency" and "structural validity" were considered not relevant for a physical performance instrument such as the 6MWT. The box "Interpretability" did not generate additional information and was excluded from analysis. The box "Generalizability" was used to evaluate general requirements: (1) whether the studies adequately described their samples in terms of age, sex, disease characteristics, setting, country, and language, and (2) whether they used adequate selection procedures and acceptable missing response rates were applied. The box "Item Response Theory" was not applicable for the studies on the evaluation of the 6MWT.

The items of each box were rated with a 4-point scoring system; excellent, good, fair, and poor. In line with the COSMIN checklist guidelines, a quality score per measurement property was obtained by taking the lowest rating of any item in a box ("worst score counts").[19]

In every box, there is 1 item that concerns the sample size requirements. The minimal requirements for an adequate sample size (N=30) as stated by the COSMIN guidelines, however, were originally developed for questionnaires and do not necessarily apply to clinimetric studies on performance measures such as the 6MWT. Clinimetric studies on performance measures tend to generate larger effect sizes and, therefore, may be appropriately evaluated using smaller sample sizes. Therefore, sample size requirements were omitted from the scoring procedure and taken into account in the best evidence synthesis instead.

Assessment of the quality criteria of the measurement properties (Tab. 2). One reviewer (B.B.) evaluated the quality of the measurement properties of the 6MWT using conventional quality criteria. The possible ratings for a measurement property were "positive," "indeterminate," and "negative."[20]

Best evidence synthesis (Tab. 3). A best evidence synthesis was performed for each patient group and based on the methodological quality

Table 3.
Best Evidence Synthesis[a]

Methodological Quality/Quality Criteria						
Diagnosis	**Study**	**Reliability**	**Agreement**	**Hypothesis Testing**	**Criterion Validity**	**Responsiveness**
Cystic fibrosis	Gulmans et al[22]	Fair/positive		Fair/positive	Excellent/positive	
	Mandrusiak et al[24]	Fair/positive				
Level of evidence		Moderate		Limited	Limited	
Cerebral palsy	Maher et al[26]	Fair/positive	Fair/indeterminate			
	Thompson et al[27]	Excellent/positive	Excellent/indeterminate			
Level of evidence		Strong	Unknown			
Duchenne muscular dystrophy	Mazzone et al[36]					Fair/positive
	McDonald et al[38]	Fair/positive	Fair/indeterminate	Fair/positive		
Level of evidence		Limited	Unknown	Limited		Limited
Spina bifida	de Groot et al[35]	Good/positive	Good/indeterminate			
Level of evidence		Limited moderate	Unknown			
Obesity	Elloumi et al[29]			Fair/positive		
	Makni et al[30]			Good/positive		
	Morinder et al[31]	Good/positive	Good/indeterminate		Fair/negative	
Level of evidence		Moderate	Unknown	Moderate	Limited	
Congenital heart disease	Moalla et al[39]		Fair/indeterminate		Excellent/positive	
Level of evidence			Unknown	Unknown	Limited	
Pulmonary hypertension	Lammers et al[42]			Fair/positive	Excellent/negative	
Level of evidence				Limited	Strong	
Juvenile idiopathic arthritis	Lelieveld et al[33]			Fair/negative	Fair/negative	
	Paap et al[34]				Fair/negative	
Level of evidence				Limited	Moderate	
End-stage renal disease	Takken et al[43]			Fair/positive	Excellent/negative	
Level of evidence				Limited	Limited	

[a] Level of evidence: strong=consistent findings in multiple studies of good methodological quality or in one study of excellent methodological quality, moderate=consistent findings in multiple studies of fair methodological quality or in one study of good methodological quality, limited=one study of fair methodological quality, conflicting=conflicting findings, unknown=only studies of poor methodological quality.

of the studies and the quality of the measurement properties.[21] Only studies with fair, good, or excellent methodological quality or generalizability were included. Separate studies on measurement properties of the same patient group were pooled. The level of evidence for each patient group was subsequently rated as "strong," "moderate," "limited," "conflicting," or "unknown." The level of evidence provided by small sample size studies (n<30) without formal power analysis was reduced to limited evidence.

Results

Included Studies

Twenty-two studies matched the inclusion criteria and were included in the systematic review (Figure). The reference checking of the included studies did not generate additional relevant studies. Reproducibility and validity of the 6MWT were evaluated in the following 11 patient groups: cystic fibrosis (n=4),[22–25] cerebral palsy (n=3),[26–28] obesity (n=4),[29–32] juvenile idiopathic arthritis (n=2),[33,34] spina bifida (n=1),[35] Duchenne muscular dystrophy (n=3),[36–38] congenital heart disease (n=1),[39] idiopathic adolescent scoliosis (n=1),[40] spinal muscular atrophy (n=1),[41] pulmonary hypertension (n=1),[42] and end-stage renal disease (n=1).[43]

Data Synthesis and Analysis

Comparison of 6MWT testing procedures (Tab. 1). Test procedures showed large variation. The walking course ranged from 8 to 70 m, but was less than 30 m in approximately half of the studies. A pretesting resting period was

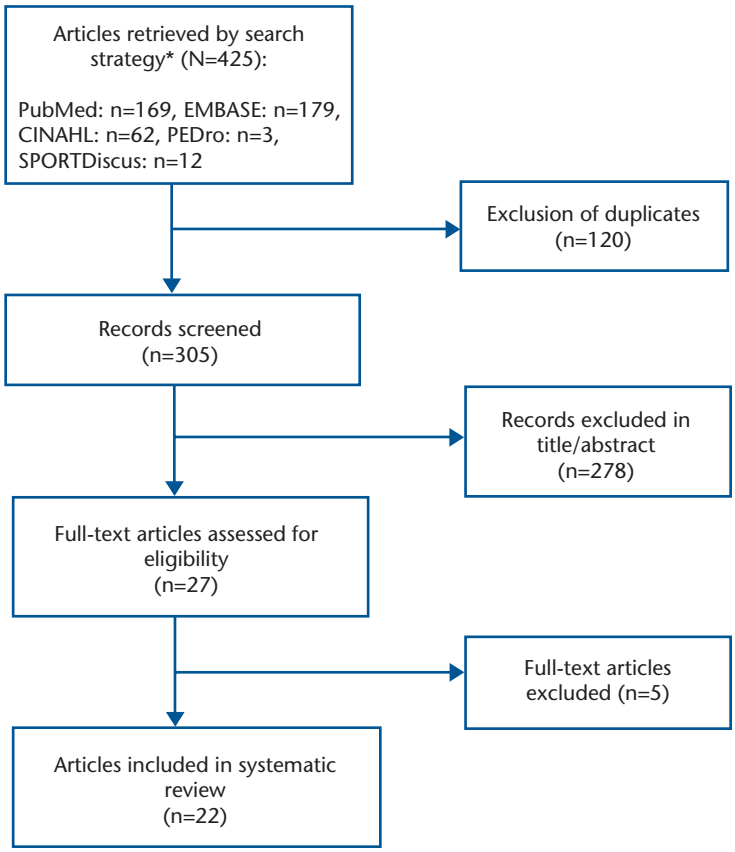

Figure.
Flow chart of the search strategy and selection of articles. Asterisk indicates February 2012.

applied in 40% of the studies. Encouragement was standardized in most of the studies. However, although 8 studies followed the standardized phrases in accordance with ATS guidelines,[17] other studies used continuous verbal encouragement. Additional procedures included the use of an instructional video,[38] visual goals for stimulating patients to keep walking,[27] and a safety chaser.[36–38] A *safety chaser* was defined as an assistant who walked behind the patient during the test and provided extra encouragement, assisted in the case of fall incidents, and performed additional measurements.[38]

Reliability (Tab. 4). Test-retest reliability was evaluated in 8 studies (see Box B: Best Evidence Synthesis of Reliability). The methodological quality of the studies on children with cystic fibrosis was rated fair[22,24] and poor.[25] The quality of the studies on children with cerebral palsy was rated excellent[27] and fair.[26] The quality of the studies on children with Duchenne muscular dystrophy,[38] obesity,[31] and spina bifida[35] was rated fair, good, and good, respectively. Main methodological flaws included inadequate statistical procedures,[25] inappropriate time intervals between tests,[24–26] and ambiguity about independent administrations.[25,26] Six studies reached high levels of test-retest reliability (ICC=.84–.98), and 1 study showed a strong correlation between the test and retest, (r_p=.90, P<.001). One study[25] was excluded from the best evidence synthesis because of statistical flaws and thus is not included in Tables 3 and 4. The level of evidence in the studies on children with Duchenne muscular dystrophy and spina bifida was reduced to limited because of their small sample sizes.

Box B: Best Evidence Synthesis of Reliability

Strong evidence was available for a positive reliability of the 6MWT in children with cerebral palsy, moderate evidence was available for a positive reliability in children with cystic fibrosis and obesity, and limited evidence was available for a positive reliability in children with Duchenne muscular dystrophy and spina bifida.

Measurement error (Tab. 4). Measurement error was evaluated in 8 studies (see Box C: Best Evidence Synthesis of Measurement Error). The methodological quality of the studies on children with cystic fibrosis was rated fair[25] and poor.[23] The quality of the studies on children with cerebral palsy was rated excellent[27] and fair.[26] The quality of the studies on children with spina bifida[35] and obesity[31] was rated good. The quality of the studies on children with congenital heart disease[39] and Duchenne muscular dystrophy[38] was rated fair and poor. Main methodological flaws were limited to incorrect calculation of the LoA in 1 study.[38] Consequently, LoA was recalculated based on the data presented. Limits of agreement were determined in 7 studies and varied from −133 m to 101 m in children with cystic fibrosis[25] and from −14 m to 12 m in children with congenital heart disease.[39] Three studies determined the SDC,[27,31] which varied between 36 m in children with spina bifida and 68 m in children with obesity.[31] It was not possible to qualify the amount of measurement error because none of the studies reported an MIC. Consequently, all 8 studies received an "indeterminate" quality rating for measurement error. One study[23] was excluded from best evidence synthesis because of inadequate description of study sample and thus is not included in Tables 3 and 4.

Table 4.
Summary of the Measurement Properties of the Studies Included for Best Evidence Synthesis: Reliability and Measurement Error[a]

Study	Population Patients (N) Age (y), $\bar{X}$(SD) Sex (M/F) Disease Characteristics	Reliability	Measurement Error 6MWD: LOA (m), SDD (m), SEM (m)
Cunha et al[25]	Cystic fibrosis (N=16) Age: 11.0 (1.9) Sex: 5/11 FEV_1 (%), $\bar{X}$ (SD) [range]: 63.1 (21.1) [30–94]		LoA: −132.7 m to 100.9 m
de Groot et al[35]	Spina bifida (N=23) Age: 10.7 (3.5) Sex: 11/12 Normal or community ambulatory	6MWD: ICC=.98 Heart rate 6MWT: ICC=.94	SEM: 13.1 m SDC: 36.3 m
Gulmans et al[22]	Cystic fibrosis (N=23) Age: 11.1 (2.2) Sex: 12/11 FEV_1 (%), $\bar{X}$(SD) [range]: 94.4 (16.5) [61–130]	r_p=.90, P<.0001	
Maher et al[26]	Cerebral palsy (N=41) Age: 13.6 (1.6) Sex: 26/15 GMFCS levels: 1–3	ICC=.98	LoA: −44 m to 42.3 m
Mandrusiak et al[24]	Cystic fibrosis (N=16) Age: 13.1 (2.7) Sex: 8/8 FEV_1 (%), $\bar{X}$ (SD) [range]: 65 (18) [36–92]	6MWD: ICC=.93 Borg Scale of Perceived Breathlessness: ICC=.92 15 c (15-count breathlessness score): ICC=.66 Spo_2: ICC=.81 Heart rate: ICC=.87	
McDonald et al[38]	DMD (N=21) Age: 8 (median), 5–12 (range) Sex: 21/0 Ambulatory (>10 m without AD)	6MWD: ICC=.91	LoA: −66 m to 74 m
Moalla et al[39]	Congenital heart disease (N=17) Age: 13.5 (0.5) Sex: ? NYHA class 2–3		LoA: −14.2 m to 11.6 m
Morinder et al[31]	Obesity (N=49) Age: 13.2 (?), 8–16 Sex: 30/19 BMI (kg/m^2): 33.9 (median), 23.3–57 (range)	ICC=.84	LoA: −65 m to 71 m SEM: 24 m SDC: 68 m
Thompson et al[27]	Cerebral palsy (N=31) Age: 9 (3) Sex: 15/16	ICC=.98 GMFCS level 1(n=9): ICC=.93; GMFCS level 2 (n=8): ICC=.91; GMFCS level 3 (n=13): ICC=.98	LoA: −71.6 to 57 m SEM: 19.8 m SDC: 54.9 m

[a] FEV_1=forced expiration volume in 1 second, GMFCS=Gross Motor Function Classification System, DMD=Duchenne muscular dystrophy, Spo_2=oxygen saturation, AD=assistive device, NYHA class=New York Heart Association Classification for heart failure, BMI=body mass index, 6MWD=6-minute walking distance, 6MWT=Six-Minute Walk Test, LoA=limits of agreement, SEM=standard error of measurement, SDC=smallest detectable change, ICC=intraclass correlation coefficient, M=male, F=female.

Box C: Best Evidence Synthesis of Measurement Error

Knowledge was available about measurement error parameters in children with cystic fibrosis, spina bifida, cerebral palsy, Duchenne muscular dystrophy, congenital heart disease, and obesity. However, the level of evidence remained unclear according to COSMIN guidelines because no information was available regarding the MIC.

Hypothesis testing (validity) (Tab. 5). Hypothesis testing was performed in 14 studies (see Box F: Best Evidence Synthesis of Hypothesis Testing), convergent validity was assessed in 7 studies, and discriminative validity was assessed in 8 studies. The methodological quality of the studies regarding children with obesity was rated fair[29,30] and poor.[31,32] The quality of the studies on children with cystic fibrosis was rated fair[22] and poor.[25] The quality of the studies on children with Duchenne muscular dystrophy,[38] pulmonary hypertension,[42] juvenile idiopathic arthritis,[33] and end-stage renal disease[43] was rated fair. The quality of the studies on children with cerebral palsy,[28] congenital heart dis-

Table 5.
Summary of the Measurement Properties of the Studies Included for Best Evidence Synthesis: Validity[a]

Study	Population Patients (N) Age (y), $\bar{X}$(SD) Sex (M/F) Disease Characteristics	Hypothesis Testing (Construct Validity)	Criterion Validity
Elloumi et al[29]	Obesity (N=28) Age: 13? Sex: 28/0 BMI (kg/m^2): >97th percentile	Convergent validity: 6MWD, BMI: r_s=−.69, P<.01, LIPOXmax: r=.67, P<.01	
Gulmans et al[22]	Cystic fibrosis (N=15) Age: 14.5 (2) Sex: 9/6 FEV_1 (%), $\bar{X}$ (SD) [range]: 58 (16.0) [41–89]	Convergent validity: 6MWD, RV/TLC (%): r_p= .72, P<.01, FEV_1(%), body weight z score, age: NS	6MWD, Wmax: r_p=.76, P<.001, $\dot{V}O_2$max: r_p=.76, P<.001, $\dot{V}O_2$max/kg: NS
Lammers et al[42]	Pulmonary hypertension (N=41) Age: 13.5 (3) Sex: 21/20 WHO class 2/3: 21/20; subgroups: IPAH (n=15), PAH+CHD (n=18), Eisenmenger syndrome (n=8)	Discriminative validity: 6MWD, −48% (17%) of predicted, P<.0001 6MWD subgroups: NS	6MWD, $\dot{V}O_2$VAT: r=.40, P=.01, $\dot{V}O_2$peak: r=.49, P=.001, $\dot{V}E/\dot{V}CO_2$: r=−.43, P=.005 Regression analysis: 6MWD (≤300 m), $\dot{V}O_2$peak: R^2=.71, P=.0006, $\dot{V}O_2$VAT: R^2=.34, P=.046, $\dot{V}E/\dot{V}CO2$: R^2=.41, P=.02 6MWD (>300 m), $\dot{V}O_2$, $\dot{V}O_2$VAT, $\dot{V}E/\dot{V}CO_2$: NS
Lelieveld et al[33]	Juvenile idiopathic arthritis Sample 1 (n=22) Age: 8.2 (2.1) Sex: 6/16 OA-JIA (n=7), Poly-JIA (n=12), S-JIA (n=3) Sample 2 (n=21) Age: 8.6 (2.0) Sex: 5/16 OA-JIA (n=12), Poly-JIA (n=9)	Discriminative validity: 6MWD lower extremity score 0–1 vs lower extremity score ≥2: P=.86	Sample 1: 6MWD, $\dot{V}O_2$peak: r_p=.43, P=.05, $\dot{V}O_2$peak/kg: NS Sample 2: 6MWD, $\dot{V}O_2$peak: r_p=.51, P=.02, 6MWD, $\dot{V}O_2$peak/kg: NS
Makni et al[30]	Obesity (N=131) Boys (n=68) Age: 13.4 (1.1) BMI (kg/m^2): 29.1 (3.3), BMI z score: 3.4 (0.7) Girls (n=63) Age: 13.8 (0.9) BMI (kg/m^2): 31 (4), BMI z score: 3.8 (0.9)	Convergent validity Boys, 6MWD, FATmax: r=.88, P<.001 Girls, 6MWD, FATmax: r=.81, P<.001	
Mazzone et al[36]	DMD (N=106) Age: 8.3 (2.3) Ambulatory (>75 m without AD), nonsteroids (n=10), intermittent (n=41), continuous (n=55)	Responsiveness: 6MWD, North Star Ambulatory Assessment: r=.52, P<.001	
McDonald et al[38]	DMD (N=21) Age: 8 (median), 5–12 (range) Sex: 21/0 Ambulatory (>10 m without AD) Control group (n=34)	Discriminative validity: DMD vs control, 6MWD (m), velocity (m/min), stride length (m): P<.001, cadence (steps/min): NS (unpaired t test) Change exercise vs resting value, DMD vs control: mean pulse rate: P=.001, mean systolic blood pressure: P=.00096, mean diastolic blood pressure: NS Convergent validity: 6MWD, age: r=.7, P=?, height, r=−0.3, P=?, stride length: r=.7, P=?, cadence: r=.8, P=?	

(Continued)

Table 5.
Continued

Study	Population Patients (N) Age (y), $\bar{X}$(SD) Sex (M/F) Disease Characteristics	Hypothesis Testing (Construct Validity)	Criterion Validity
Moalla et al[39]	Congenital heart disease (N=17) Age: 13.5 (0.5) Sex: ? NYHA class 2–3, control group (n=12)		6MWD, $\dot{V}o_2$max: r=.76, P<.01, $\dot{V}o_2Vt$, r=.69, P<.05
Morinder et al[31]	Obesity (N=250) Age: 13.3 (median), 8–16 (range) Sex: 126/124 BMI (kg/m^2): 35.9 (5.5)		6MWD, $\dot{V}o_2$max estimated: r=.34, P<.001
Paap et al[34]	Juvenile idiopathic arthritis (N=18) Age: 13.3 (3.1) Sex: 6/12 OA-JIA (n=7), Poly-JIA (n=7), S-JIA (n=3), PsA (n=1)		6MWD, $\dot{V}o_2$peak: r=.53, P=.02
Takken et al[43]	End-stage renal disease (N=20) Age: 14.1 (3.4) Sex: 13/7 Time on dialysis: 27.5 (38.2)	Discriminative validity: 6MWD reference values: P<.0001 Convergent validity: 6MWD, hematocrit: r=.7, hemoglobin: r=.7, VT: r=?, height: r=? (P<.05); 6MWD muscle strength, weight, BMI: NS Regression model: R^2=.60, hematocrit; β coefficient=1.8	6MWD, $\dot{V}o_2$peak: NS

[a] WHO=World Health Organization, IPAH=idiopathic pulmonary hypertension, OA-JIA=oligoarticular juvenile idiopathic arthritis, Poly-JIA=polyarticular juvenile idiopathic arthritis, S-JIA=systemic juvenile idiopathic arthritis, NYHA class=New York Heart Association Classification for heart failure, PsA=psoriatic arthritis, BMI=body mass index, 6MWD=6-minute walk distance, AD=assistive device, LIPOXmax=maximal lipod oxidation, FEV_1=forced expiration volume in 1 second, RV/TLC=residual volume/total lung capacity ratio, NS=nonsignificant, $\dot{V}o_2$peak=peak oxygen uptake, $\dot{V}o_2$max=maximum oxygen uptake, M=male, F=female, PAH+CHD=pulmonary hypertension associated with congenital heart defects, Wmax=maximum workload, $\dot{V}o_2$VAT=oxygen uptake at anaerobic threshold, $\dot{V}E/\dot{V}co_2$=ventilator efficiency, $\dot{V}o_2$Vt=oxygen uptake at ventilator threshold, FATmax=maximal fat oxidation rate.

ease,[39] idiopathic adolescent scoliosis,[40] and spinal muscular atrophy[41] was rated poor. Main methodological flaws included the absence of a predefined hypothesis,[28,32,39–41] inadequate statistical analysis,[25,39] and validity issues concerning test performance.[38] Seven studies[25,28,31,32,39–41] were excluded from best evidence synthesis because of major methodological flaws and thus are not included in Tables 3 and 5.

Box F: Best Evidence Synthesis of Hypothesis Testing

There was moderate evidence for a positive convergent validity of the 6MWT with fat oxidation in children with obesity. Limited evidence was available for: (1) positive convergent validity of the 6MWT with disease specific parameters in children with cystic fibrosis (pulmonary hyperinflation) and with end-stage renal disease (hematocrit); (2) good discriminative validity with control groups or subgroups in children with Duchenne muscular dystrophy, end-stage renal disease, and pulmonary hypertension; and (3) poor discriminative validity between subgroups of children with juvenile idiopathic arthritis.

Criterion validity (validity) (Tab. 5). Criterion validity was evaluated in 8 studies (see Box H: Best Evidence Synthesis of Criterion Validity). In all of the studies, the 6MWT was compared with peak oxygen uptake ($\dot{V}o_2$peak) using an incremental cycle ergometer test, 6 studies used a maximal protocol to ensure true $\dot{V}o_2$peak, and 2 studies predicted $\dot{V}o_2$peak based on a submaximal protocol. The methodological quality of the studies on children with cystic fibrosis,[22] congenital heart disease,[39] pulmonary hypertension,[42] and end-stage renal disease[43] was rated excellent. The quality of the studies on children with obesity was rated fair[31] and poor.[29] The quality of the studies on children with juvenile idiopathic arthritis was rated fair.[33,34] Main methodological flaws included bias related to magnitude of the scores.[29] High significant correlations between the 6MWD and $\dot{V}o_2$peak were found in children

with cystic fibrosis (r=.76)[22] and congenital heart disease (r=.76)[39] and between the 6MWD and maximum oxygen uptake according to American College of Sports Medicine guideline in children with obesity (r=.77).[29] Weak correlations were found in children with pulmonary hypertension (r=.49),[42] juvenile idiopathic arthritis (r=.43–.53),[33,34] obesity (r=.34),[31] and end-stage renal disease (not significant).[43] One study[29] was excluded from best evidence synthesis because of statistical flaws and thus is not included in Tables 3 and 5. The level of evidence in the studies on children with cystic fibrosis, congenital heart disease, and end-stage renal disease was reduced to limited because of the small sample sizes.

Box H: Best Evidence Synthesis of Criterion Validity

There was strong evidence for poor criterion validity of the 6MWT in children with pulmonary hypertension. Moderate evidence was available for poor criterion validity in children with juvenile idiopathic arthritis. There was limited evidence for: (1) good criterion validity in children with cystic fibrosis and congenital heart disease and (2) poor criterion validity in children with obesity and end-stage renal disease.

Responsiveness (Tab. 5). Responsiveness was evaluated in 2 studies, both including children with Duchenne muscular dystrophy (see Box I: Best Evidence Synthesis of Responsiveness). The methodological quality of the studies was rated fair[36] and poor.[37] Main methodological flaws included the absence of predefined hypotheses about change scores,[36,37] absence of a comparator instrument,[37] and inadequate description of the interim period.[36,37] In 1 study,[36] a moderate correlation was found between change scores on the 6MWT and another functional walking test (r=.52, P<.001). One study[37] was excluded from best evidence synthesis because of methodological flaws and thus is not included in Tables 3 and 5.

Box I: Best Evidence Synthesis of Responsiveness

Limited evidence was available for a positive responsiveness of the 6MWT in children with Duchenne muscular dystrophy.

Discussion

Twenty-two studies were evaluated on both methodological quality and quality criteria of the measurement properties of the 6MWT in chronic pediatric conditions. Seven studies showed poor methodological quality or poor generalizability and were excluded from best evidence synthesis. The best evidence synthesis of the 15 included studies provides an overview of the current body of knowledge about the measurement properties of the 6MWT in chronic pediatric conditions.

Methodological Considerations

The methodological quality of the studies varied between poor and excellent. Notably, the studies on hypothesis testing received remarkably low ratings on methodological quality. These low ratings were due mainly to the fact that few studies formulated clearly defined hypotheses. Without specific hypotheses on expected mean differences between known groups or expected correlations with other variables, it remains unclear whether the results reflect the construct to be measured and only little can be said about the validity of the 6MWT.[44]

The importance of a clear statement on expected correlations in the assessment of validity of the 6MWT is even more important if we consider that the construct to be measured with the 6MWT in a given population is not clearly described in most of the studies. Moreover, the particular construct of the 6MWT in each case seems to depend largely on both the origin and severity of the functional limitations. The 6MWT reflects maximal exercise capacity in pediatric patients with moderate to severe pulmonary and cardiovascular conditions such as cystic fibrosis,[22] congenital heart disease,[39] and severe pulmonary hypertension,[42] but submaximal exercise capacity in other chronic conditions.[31,33,34] This finding is in accordance with a prior study regarding the adult population.[45] Jehn et al[45] showed that the 6MWT reflects a maximum exercise response in patients with advanced heart failure, whereas it only constitutes a submaximal exercise test in patients with mild heart failure and no functional limitations. Given the fact that the construct of the 6MWT depends on both the patient population and the disease severity, it seems no longer justified to label the 6MWT as a submaximal functional outcome measure before a proper validity assessment on the target population is performed, including both patients who are mildly and severely affected.

Evaluative Properties of the 6MWT (Reliability and Measurement Error)

The reproducibility of an instrument reflects both reliability and measurement error parameters. Whereas the first parameter refers to the ability to distinguish among individuals and is mostly important when used for discriminative purposes, the latter parameter refers to the ability of the instrument to detect relevant changes and is more suitable for evaluative purposes. The present systematic review provides sufficient evidence that the 6MWT is capable of distinguishing not only between children with chronic conditions and their peers who are healthy but also within patient populations. Although the measurement error for several subgroups has been

determined, the evaluative value of the 6MWT remains unclear. All studies on measurement error received an "indeterminate" rating because none of them calculated or reported an MIC. This rating seems rather strict considering that the measurement error of the 6MWT has been defined in various patient groups. However, these parameters can be qualified only in relation to the MIC. Without knowledge about the MIC, it is difficult to interpret the observed changes and determine the success of an intervention. The MIC of the 6MWT has been investigated in recent clinimetric studies on adult patients with chronic conditions, resulting in an MIC of 24 to 45 m.[46,47] Future studies on the 6MWT in the pediatric population should follow this initiative and include the assessment of the MIC in their clinimetric evaluation of the 6MWT.

Sample Size

The sample size in most of the included studies was small. In general, small sample sizes lead to reduction of power and hinder the ability to generalize the results to the reference population.[48] However, the required sample size to generate sufficient power in a study is not fixed but can be calculated based on the effect size and the significance criterion. Regrettably, none of the studies reported a power analysis; therefore, it remained unclear whether the sample sizes were adequate. In the best evidence synthesis, studies on the same patient group with sufficient methodological quality were pooled to increase sample size and extend the level of evidence. A formal meta-analysis with the COSMIN, which might increase power sufficiently to detect important differences, was not possible. The use of this technique in systematic reviews evaluated with the COSMIN is under development.

Test Procedures

There was a large variation in test procedures among the included studies, and only 1 study followed all ATS guidelines. These differences in length of the walking course,[49] the means of encouragement,[50] and the use of a practice test and resting period might have negatively influenced the generalizability of the results to the reference population and limited the accuracy of the calculated measurement error within the studies. It is important, therefore, that the same standardized guidelines are used in all studies of the 6MWT. Given the variability among the studies, it is important to consider whether modifications should be made for the pediatric population. Children tend to have more difficulty performing tasks that take longer and possibly perform better with more frequent verbal and visual encouragement. However, the need for continuous encouragement is in contrast to the concept of a self-paced walking test and might change the construct validity of the test. It is possible that the variability in walking courses can be explained by the difference in the length of available straight corridors among test locations. The influence of the length of the walking course is small, provided that it is a straight path of at least 15 m long. Therefore, to enable future researchers and clinical practitioners to follow existing guidelines and improve homogeneity of test procedures, it might be useful to reduce the current criterion of 30 m (as stated in the ATS guidelines) to 15 m.

Limitations of the Review

A specific checklist to evaluate the methodological quality of clinimetric studies regarding performance measures was not available. The COSMIN was originally developed for health-related, patient-reported outcome measures, such as questionnaires. Therefore, the validity and reliability of the COSMIN itself, as a tool for assessing the methodological quality of studies examining performance tests such as the 6MWT, can be questioned. The boxes, internal consistency, and structural validity turned out to be irrelevant in the methodological assessment of the studies on the 6MWT, and several items such as "highest and lowest possible scores" were scored as not applicable. Nonetheless, in the absence of a specific measure specifically designed to evaluate clinimetric studies, the COSMIN seems to be a good alternative tool for gaining more insight into the methodological quality of studies on performance tests, in addition to serving as a useful checklist for establishing methodologically proper research protocols for clinimetric studies.

Studies that were not available in full text or were published in a language other than English or Dutch were excluded. As a result, some information on the measurement properties of the 6MWT might have been missed that could have increased the level of evidence for specific populations.

Clinical Implications

Clinicians can use the results of this study to support their diagnostic process and evaluation of interventions in children with chronic conditions. The best evidence synthesis provides a clear overview of the current knowledge of the reliability and validity of the 6MWT in specific chronic conditions and can guide clinicians in their decision of whether to use the 6MWT and in interpreting the outcome. If a clinician, for example, wants to evaluate physical fitness in a child with juvenile idiopathic arthritis, he or she might consider the 6MWT. The best evidence synthesis of studies of juvenile idiopathic arthritis, however, shows that the 6MWT is likely to be a poor indicator for aerobic capacity and

that it is unknown whether the 6MWT provides reproducible test results. Based on this information, a clinician will possibly decide to use another outcome measure such as the maximal cardiopulmonary exercise test. The lack of evidence for the measurement properties of the 6MWT for many pediatric chronic conditions does not automatically mean that the 6MWT cannot be applied. It does demand a critical approach of the clinician to the aim and the results of the test. Without knowledge about the construct of the 6MWT for a specific population and the measurement error between repeated measurements, results should be interpreted with caution and alternative outcome measures must be considered.

Conclusions

Evidence for measurement properties of the 6MWT varies greatly among chronic pediatric conditions. Further research is needed in all patient groups to explore the ability of the 6MWT to measure significant and clinically important changes. Until further research is conducted, changes measured with the 6MWT should be interpreted with caution and attention should be paid to whether it is a meaningful change for the individual patient. Future studies toward consensus regarding modified test procedures in the pediatric population are recommended.

All authors provided concept/idea/research design, writing, and data analysis. Mr Bartels and Dr de Groot provided data collection. Mr Bartels provided project management and facilities/equipment. Dr de Groot and Dr Terwee provided consultation (including review of manuscript before submission).

This study was part of the Healthy Active Living in Youth With Neuromotor Disability (HALYNeD) study and was partly financially supported by SIA RAAK Publiek.

DOI: 10.2522/ptj.20120210

References

1 ATS Committee on Proficiency Standards for Clinical Pulmonary Function Laboratories. ATS statement: guidelines for the Six-Minute Walk Test. *Am J Respir Crit Care Med.* 2002;166:111–117.

2 Douma-van Riet DC, Engelbert RH, van Genderen FR, et al. Physical fitness in children with haemophilia and the effect of overweight. *Haemophilia.* 2009;15:519–527.

3 Vis JC, Thoonsen H, Duffels MG, et al. Six-minute walk test in patients with Down syndrome: validity and reproducibility. *Arch Phys Med Rehabil.* 2009;90:1423–1427.

4 Kierkegaard M, Harms-Ringdahl K, Edstrom L, et al. Feasibility and effects of a physical exercise programme in adults with myotonic dystrophy type 1: a randomized controlled pilot study. *J Rehabil Med.* 2011;43:695–702.

5 Calders P, Deforche B, Verschelde S, et al. Predictors of 6-minute walk test and 12-minute walk/run test in obese children and adolescents. *Eur J Pediatr.* 2008;167:563–568.

6 Hanewinkel-van Kleef YB, Helders PJ, Takken T, Engelbert RH. Motor performance in children with generalized hypermobility: the influence of muscle strength and exercise capacity. *Pediatr Phys Ther.* 2009;21:194–200.

7 Wraith JE, Clarke LA, Beck M, et al. Enzyme replacement therapy for mucopolysaccharidosis I: a randomized, double-blinded, placebo-controlled, multinational study of recombinant human alpha-L-iduronidase (laronidase). *J Pediatr.* 2004;144:581–588.

8 Goemans NM, Tulinius M, Van Den Akker JT, et al. Systemic administration of PRO051 in Duchenne's muscular dystrophy [erratum in: *N Engl J Med*. 2011;365:1361]. *N Engl J Med.* 2011;364:1513–1522.

9 de Groot JF, Takken T, van Brussel M, et al. Randomized controlled study of home-based treadmill training for ambulatory children with spina bifida. *Neurorehabil Neural Repair.* 2011;25:597–606.

10 Geiger R, Strasak A, Treml B, et al. Six-minute walk test in children and adolescents. *J Pediatr.* 2007;150:395–399, e1–e2.

11 Bushby K, Finkel RS, Birnkrant DJ, et al. Diagnosis and management of Duchenne muscular dystrophy, part 2: implementation of multidisciplinary care [erratum in: *Lancet Neurol*. 2010;9:237]. *Lancet Neurol.* 2010;9:177–189.

12 de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol.* 2006;59:1033–1039.

13 Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol.* 2010;63:737–745.

14 Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol.* 2003;56:395–407.

15 Terwee CB, Jansma EP, Riphagen II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res.* 2009;18:1115–1123.

16 Ebbeling CB, Pawlak DB, Ludwig DS. Childhood obesity: public-health crisis, common sense cure. *Lancet.* 2002;360:473–482.

17 ATS Committee on Proficiency Standards for Clinical Pulmonary Function Laboratories. ATS statement: guidelines for the Six-Minute Walk Test. *Am J Respir Crit Care Med.* 2002:111–117.

18 The COSMIN checklist. EMGO Institute for Health and Care Research, VU University Medical Center. Available at: http://www.cosmin.nl/the-cosmin-checklist_8_5.html.

19 Terwee CB, Mokkink LB, Knol DL, et al. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res.* 2012;21:651–657.

20 Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* 2007;60:34–42.

21 Schellingerhout JM, Verhagen AP, Heymans MW, et al. Measurement properties of disease-specific questionnaires in patients with neck pain: a systematic review. *Qual Life Res.* 2012;21:659–670.

22 Gulmans VA, van Veldhoven NH, de Meer K, Helders PJ. The Six-Minute Walking Test in children with cystic fibrosis: reliability and validity. *Pediatr Pulmonol.* 1996;22:85–89.

23 Balfour-Lynn IM, Prasad SA, Laverty A, et al. A step in the right direction: assessing exercise tolerance in cystic fibrosis. *Pediatr Pulmonol.* 1998;25:278–284.

24 Mandrusiak A, Maurer C, MacDonald J, et al. Functional capacity tests in young people with cystic fibrosis. *NZ J Physiother.* 2009;37:13–16.

25 Cunha MT, Rozov T, de Oliveira RC, Jardim JR. Six-minute walk test in children and adolescents with cystic fibrosis. *Pediatr Pulmonol.* 2006;41:618–622.

26 Maher CA, Williams MT, Olds TS. The six-minute walk test for children with cerebral palsy. *Int J Rehabil Res.* 2008;31:185–188.

27 Thompson P, Beath T, Bell J, et al. Test-retest reliability of the 10-metre fast walk test and 6-minute walk test in ambulatory school-aged children with cerebral palsy. *Dev Med Child Neurol.* 2008;50:370–376.

28 Chong J, Mackey AH, Broadbent E, Stott S. Relationship between walk tests and parental reports of walking abilities in children with cerebral palsy. *Arch Phys Med Rehabil.* 2011;92:265–270.

29 Elloumi M, Makni E, Ounis OB, et al. Six-minute walking test and the assessment of cardiorespiratory responses during weight-loss programmes in obese children. *Physiother Res Int.* 2011;16:32–42.

30 Makni E, Moalla W, Trabelsi Y, et al. Six-minute walking test predicts maximal fat oxidation in obese children. *Int J Obes (Lond).* 2012;36:908–913.

31 Morinder G, Mattsson E, Sollander C, et al. Six-minute walk test in obese children and adolescents: reproducibility and validity. *Physiother Res Int.* 2009;14:91–104.

32 Guinhouya B. Outcomes and cardiac response of overweight prepubescent to the 6 minutes walk test. *Minerva Pediatr.* 2011;63:375–384.

33 Lelieveld OT, Takken T, van der Net J, van Weert E. Validity of the 6-minute walking test in juvenile idiopathic arthritis. *Arthritis Rheum.* 2005;53:304–307.

34 Paap E, van der Net J, Helders PJ, Takken T. Physiologic response of the six-minute walk test in children with juvenile idiopathic arthritis. *Arthritis Rheum.* 2005;53:351–356.

35 de Groot JF, Takken T, Gooskens RH, et al. Reproducibility of maximal and submaximal exercise testing in "normal ambulatory" and "community ambulatory" children and adolescents with spina bifida: which is best for the evaluation and application of exercise testing? *Phys Ther.* 2011;91:267–276.

36 Mazzone E, Vasco G, Sormani MP, et al. Functional changes in Duchenne muscular dystrophy: a 12-month longitudinal cohort study. *Neurology.* 2011;77:250–256.

37 McDonald CM, Henricson EK, Han JJ, et al. The 6-minute walk test in Duchenne/Becker muscular dystrophy: longitudinal observations. *Muscle Nerve.* 2010;42:966–974.

38 McDonald CM, Henricson EK, Han JJ, et al. The 6-minute walk test as a new outcome measure in Duchenne muscular dystrophy. *Muscle Nerve.* 2010;41:500–510.

39 Moalla W, Gauthier R, Maingourd Y, Ahmaidi S. Six-minute walking test to assess exercise tolerance and cardiorespiratory responses during training program in children with congenital heart disease. *Int J Sports Med.* 2005;26:756–762.

40 Alves VL, Avanzi O. Objective assessment of the cardiorespiratory function of adolescents with idiopathic scoliosis through the Six-Minute Walk Test. *Spine (Phila Pa 1976).* 2009;34:926–929.

41 Montes J, McDermott MP, Martens WB, et al. Six-Minute Walk Test demonstrates motor fatigue in spinal muscular atrophy. *Neurology.* 2010;75:833–838.

42 Lammers AE, Diller GP, Odendaal D, et al. Comparison of 6-min walk test distance and cardiopulmonary exercise test performance in children with pulmonary hypertension. *Arch Dis Child.* 2011;96:141–147.

43 Takken T, Engelbert R, van Bergen M, et al. Six-minute walking test in children with ESRD: discrimination validity and construct validity. *Pediatr Nephrol.* 2009;24:2217–2223.

44 Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol.* 2010;10:22.

45 Jehn M, Halle M, Schuster T, et al. The 6-min walk test in heart failure: is it a max or sub-maximum exercise test? *Eur J Appl Physiol.* 2009;107:317–323.

46 du Bois RM, Weycker D, Albera C, et al. Six-minute-walk test in idiopathic pulmonary fibrosis: test validation and minimal clinically important difference. *Am J Respir Crit Care Med.* 2011;183:1231–1237.

47 Gremeaux V, Troisgros O, Benaim S, et al. Determining the minimal clinically important difference for the six-minute walk test and the 200-meter fast-walk test during cardiac rehabilitation program in coronary artery disease patients after acute coronary syndrome. *Arch Phys Med Rehabil.* 2011;92:611–619.

48 Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Practice.* 3rd ed. Upper Saddle River, NJ: Pearson Prentice Hall; 2008.

49 Weiss RA. Six minute walk test in severe COPD: reliability and effect of walking course layout and length. Paper presented at: ACCP Conference; September 2000; San Francisco, California.

50 Guyatt GH, Pugsley SO, Sullivan MJ, et al. Effect of encouragement on walking test performance. *Thorax.* 1984;39:818–822.