

M.A.T. Bloemen, MSc, Research Group Lifestyle and Health, HU University of Applied Sciences Utrecht, Bolognalaan 101, PO Box 85182 3508 AD, Utrecht, the Netherlands; Child Development and Exercise Center, Wilhelmina Children's Hospital, University Medical Center Utrecht, Utrecht, the Netherlands; and Department of Rehabilitation, Nursing Science and Sports, University Medical Center Utrecht, Utrecht, the Netherlands. Address all correspondence to Ms Bloemen at: manon.bloemen@hu.nl.

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

F.J.G. Backx, MD, PhD, Department of Rehabilitation, Nursing Science and Sports, University Medical Center Utrecht.

J. Benner, MSc, Research Group Lifestyle and Health, HU University of Applied Sciences Utrecht.

C.L.J.J. Kruitwagen, MSc, Department of Biostatistics and Research Support, Julius Center, University Medical Center Utrecht.

T. Takken, PhD, Child Development and Exercise Center, Wilhelmina Children's Hospital, University Medical Center Utrecht.

[Bloemen MAT, Groot JFD, Frank J.G. Backx FJG, et al. Wheelchair shuttle test for assessing aerobic fitness in youth with spina bifida: validity and reliability. *Phys Ther*. 2017;97:1020-1029.]

© American Physical Therapy Association 2017. Published by Oxford University Press [on behalf of the American Physical Therapy Association]. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Published Ahead of Print:

August 3, 2017

Accepted: July 16, 2017

Submitted: March 8, 2016

## Wheelchair Shuttle Test for Assessing Aerobic Fitness in Youth With Spina Bifida: Validity and Reliability

Manon A.T. Bloemen, Janke F. de Groot, Frank J.G. Backx, Joyce Benner, Cas L.J.J. Kruitwagen, Tim Takken

**Background.** Testing aerobic fitness in youth is important because of expected relationships with health.

**Objective.** The purpose of the study was to estimate the validity and reliability of the Shuttle Ride Test in youth who have spina bifida and use a wheelchair for mobility and sport.

**Design.** This study is a validity and reliability study.

**Methods.** The Shuttle Ride Test, Graded Wheelchair Propulsion Test, and skill-related fitness tests were administered to 33 participants for the validity study (age =  $14.5 \pm 3.1$  y) and to 28 participants for the reliability study (age =  $14.7 \pm 3.3$  y).

**Results.** No significant differences were found between the Graded Wheelchair Propulsion Test and the Shuttle Ride Test for most cardiorespiratory responses. Correlations between the Graded Wheelchair Propulsion Test and the Shuttle Ride Test were moderate to high ( $r = .55-.97$ ). The variance in peak oxygen uptake ( $VO_{2peak}$ ) could be predicted for 77% of the participants by height, number of shuttles completed, and weight, with large prediction intervals. High correlations were found between number of shuttles completed and skill-related fitness tests ( $CI = .73$  to  $-.92$ ). Intraclass correlation coefficients were high ( $.77-.98$ ), with a smallest detectable change of 1.5 for number of shuttles completed and with coefficients of variation of 6.2% and 6.4% for absolute  $VO_{2peak}$  and relative  $VO_{2peak}$ , respectively.

**Conclusions.** When measuring  $VO_{2peak}$  directly by using a mobile gas analysis system, the Shuttle Ride Test is highly valid for testing  $VO_{2peak}$  in youth who have spina bifida and use a wheelchair for mobility and sport. The outcome measure of number of shuttles represents aerobic fitness and is also highly correlated with both anaerobic performance and agility. It is not possible to predict  $VO_{2peak}$  accurately by using the number of shuttles completed. Moreover, the Shuttle Ride Test is highly reliable in youth with spina bifida, with a good smallest detectable change for the number of shuttles completed.



Post a comment for this article at:

<https://academic.oup.com/ptj>

Testing aerobic fitness in youth with a physical disability like spina bifida is currently an important aspect of pediatric physical therapy because of the expected relationships between fitness and health.<sup>1,2</sup> The aerobic fitness of youth with spina bifida is low compared with that of peers with typical development but also compared with that of youth with other chronic diseases.<sup>3-5</sup> A study with adolescents and young adults with spina bifida showed that the average aerobic fitness was 42% lower than that of their peers with typical development, with the lowest scores in participants who use wheelchairs.<sup>6</sup>

We analyzed the measurement properties of the Graded Wheelchair Propulsion Test (GWPT), a highly valid and highly reliable laboratory test to measure peak oxygen uptake ( $VO_{2peak}$ ) in youth with spina bifida who use a wheelchair for mobility and sport.<sup>7</sup> Despite the benefits of a laboratory test, field tests are more applicable for pediatric physical therapists as these tests can be performed in their own setting without the investment of expensive and sophisticated equipment. Valid and reliable field-based tests will contribute to evaluation of interventions and the clinical reasoning process of pediatric physical therapists concerning aerobic fitness.

The Shuttle Ride Test (SRiT), derived from the well-known and frequently used Shuttle Run Test in youth who are ambulatory, has been used in other clinical populations and seems to be the most appropriate maximal aerobic field test for measuring aerobic fitness in youth with spina bifida who use a wheelchair for mobility and sport.<sup>8,9</sup> During the SRiT, which is a stepwise incremental maximal exercise test, children propel their wheelchair back and forth to exhaustion with a standardized increasing speed between 2 lines that are 10 m apart. The main outcome measure is the number of shuttles they achieved, with 1 shuttle corresponding to approximately 1 minute of wheelchair propulsion. A shuttle is a stage with a constant speed and the speed is increased approximately every 1 minute. This principle has been used over decades in field exercise tests in

children.<sup>10</sup> The SRiT is a highly valid test for measuring  $VO_{2peak}$  when using a mobile gas analysis system in youth with cerebral palsy (CP) who have Gross Motor Function Classification System scores of III and IV and who use a wheelchair for mobility and sport. As disorder-related characteristics differ between youth with CP and spina bifida, the validity should also be estimated for youth with spina bifida who use a wheelchair for mobility and sport, comparing  $VO_{2peak}$  of the SRiT to the GWPT using a mobile gas analysis system.

Although criterion validity concerns comparing  $VO_{2peak}$  measures between the SRiT and the GWPT, most pediatric physical therapists and other clinicians do not have the access to a mobile gas analysis system when assessing the SRiT. They have to use the metric shuttle during their clinical reasoning process. Evidence in athletes who use wheelchairs showed that it is difficult to predict  $VO_{2peak}$  using the distance travelled during the incremental SRiT.<sup>11,12</sup> In children with osteogenesis imperfecta, poor correlations were found between  $VO_{2peak}$  and the outcome measure shuttles.<sup>9</sup> Furthermore, the importance of anaerobic performance and agility during the SRiT has been hypothesized, as turning, deceleration, and acceleration seem to be important skills during the SRiT.<sup>8,12</sup> At the same time both personal aspects (eg, age) and wheelchair features may play an important role in the number of shuttles that a child achieves during the SRiT.<sup>12,13</sup> So even though literature suggests that the SRiT is a highly valid, inexpensive field test for measuring aerobic fitness in children with CP and osteogenesis imperfecta who use a wheelchair for mobility and sport, the construct of the metric shuttles is still unclear. Therefore, it is interesting to evaluate the possibility to predict  $VO_{2peak}$  achieved during the SRiT by using the outcome measure shuttles and to test the hypotheses of moderate to high correlations between the shuttles and anaerobic performance, agility, personal factors and wheelchair features in children with spina bifida who use a wheelchair for mobility and sport. This information will help pediatric physical therapists to interpret the outcome measure shuttles during their clinical reasoning process.

Besides validity, reliability is an important measurement property and highly relevant when evaluating the effects of training. The SRiT is highly reliable in youth with CP and osteogenesis imperfecta who use a wheelchair for mobility and sport, with intraclass correlation coefficients (ICCs) of greater than .95.<sup>8,9</sup> The reliability of the SRiT has yet to be established in youth with spina bifida who use a wheelchair for mobility and sport.

Therefore, the aims of this study are to estimate the criterion validity of the SRiT by comparing cardiorespiratory responses of the SRiT to the GWPT using a mobile gas analysis system, to estimate the construct validity of the metric shuttles by predicting  $VO_{2peak}$  and correlating the shuttles with anaerobic performance, agility, personal factors and wheelchair features, and to estimate the reliability of the SRiT in youth with spina bifida who use a wheelchair for mobility and sport.

## Methods

### Participants

This study is part of the larger “Let’s Ride...” study, looking at fitness and physical activity in youth with spina bifida who use a wheelchair for mobility and sport.<sup>7,14</sup> The recruitment of participants took place in the Netherlands by pediatric physical therapists, rehabilitation centers, the BOSK (Association of an by parents of youth and adults with a disability) and spina bifida outpatient services. Participants were included if they were diagnosed with spina bifida, 5 to 18 years of age during enrollment, and were able to follow instructions of the measurements. They had to self-propel a wheelchair during daily life, long distances or sports participation, meaning that they had to be experienced wheelchair-users. Participants were excluded if any (medical) event was present that intervened with testing outcome.

### Procedures

The study procedures were approved by the Medical Ethics Committee of the University Medical Center Utrecht, Utrecht, the Netherlands. Children aged 12 years and older and all parents were

required by Dutch law and regulations to sign informed consent forms.

The participants of the validity study were measured twice (1 day for the GWPT and 1 day for the SRiT), those who also participated in the reliability study were measured 3 times (a third day for the second SRiT), with 3 days to 1 week between testing. The laboratory test GWPT was measured in the laboratory of the HU University of Applied Sciences (Utrecht, the Netherlands). All field tests were measured either in the gymnasium of the HU University of Applied Sciences or in a gymnasium nearby the participants home. The participants were always tested in similar conditions: in their own wheelchair, with the same tire pressure (maximum allowed) and on the same floor. The testing order differed between the participants, because of practical aspects like availability of the gymnasium or laboratory. Only 1 maximal aerobic exercise test was performed on a single day and the maximal aerobic exercise test was always the last test of that day. Between the other short duration tests, a resting period of at least 5 minutes was scheduled. A standard questionnaire recorded age, sex, type of spina bifida, lesion level, use of wheelchair and type of wheelchair. Before testing, body mass and wheelchair mass were determined by an electronic wheelchair scale (Kern MWS-300K100M; Kern & Sohn GmbH, Balingen, Germany). Arm span while seated as proxy for height was measured through the use of non-stretchable tape (middle fingertip to middle fingertip) as recommended for children who use wheelchairs, because of possible contractures in hips and knees when lying supine.<sup>15</sup> The body mass index (BMI; body mass divided by the square of length) was adjusted with  $\times 0.95$  for midlumbar lesions and  $\times 0.90$  for high lumbar/thoracic lesions.<sup>15</sup>

### Maximal aerobic exercise testing.

Two graded exercise tests, the GWPT and the SRiT, were performed by all participants. A heart rate (HR) monitor (miniCardio; Hosand Technologies Srl, Verbania, Italy) measured HR during the tests and a calibrated mobile gas analysis system (Cortex Metamax

B3; Cortex Medical GmbH, Leipzig, Germany) was used to measure cardiorespiratory responses during both the GWPT and the SRiT. The Cortex Metamax has been used in multiple studies regarding exercise testing in youth who use a wheelchair for mobility and sport and is valid and reliable for measuring gas-exchange parameters during exercise.<sup>7,9,16</sup>

Both the GWPT and the SRiT were continued until the participants stopped due to exhaustion, despite verbal encouragement from the test leader (M.A.T.B.), who was an experienced pediatric physical therapist.

### Graded Wheelchair Propulsion Test (GWPT).

The incremental protocol of the GWPT is described in an earlier study, it is a highly valid and reliable test to assess  $VO_{2peak}$  (ICCs  $> .9$ ) in youth with spina bifida.<sup>7</sup> The participants were sitting in their own wheelchair and were secured to a wheelchair ergometer (custom-made, based on the model 3600 CatEye ergociser; CatEye, Osaka, Japan). The participants had to maintain their self-selected comfortable wheelchair propulsion speed (60–120 rpm), while the resistance was increased by 0.1-torque increments every minute<sup>7</sup>.

### Shuttle Ride Test (SRiT).

The protocol of the SRiT, as described earlier by Verschuren et al<sup>8</sup> in youth with CP and by Bongers et al<sup>9</sup> in youth with osteogenesis imperfecta. The participants had to propel their wheelchair back and forth from one line to another line, 10 m apart. Participants were instructed to cross the lines with their front wheels and then turn 180 degrees and proceed without stopping. The starting speed was 2.0 km/h and the speed increased with 0.25 km/h every minute. This incremental pace was dictated by an auditory cue (“beep”) played by a standard CD player, so the children knew when to start and in what pace they had to propel their wheelchair between the lines. The main performance outcome measure of the SRiT is the total number of achieved shuttles (ranging from 0.5 to 23 shuttles), with 1 shuttle corresponding to approximately 1 minute of wheelchair propulsion.

The children had to continue until they failed to reach the line within 1.5 m, on 2 consecutive beeps, despite verbal encouragements. All participants were accompanied by the test leader to help pace themselves and to encourage them to achieve maximal effort.<sup>8,9</sup> Next to cardiorespiratory responses, the total number of achieved shuttles was recorded.

### Exercise testing parameters: cardiorespiratory responses.

Objective criteria for maximal aerobic exercise testing for ambulating children were defined as a  $HR_{peak} \geq 180/min$ , peak respiratory exchange ratio ( $RER_{peak}$ ) of greater than 0.99 or the presence of a  $VO_2$  plateau.<sup>17</sup> The applicability of these objective criteria for youth who use a wheelchair is unclear, so in this study, data of both the GWPT and the SRiT were included for analysis if the subjective criteria (signs of intense effort such as sweating, facial flushing, clear unwillingness to continue despite encouragement) were met.<sup>17</sup> Original data from the Cortex Metamax were prepared for analysis using a 10-second moving average interval.<sup>18</sup> Absolute  $VO_{2peak}$ , peak carbon dioxide production ( $VCO_{2peak}$ ),  $RER_{peak}$ , peak minute ventilation ( $VE_{peak}$ ), peak tidal volume ( $TV_{peak}$ ) and peak breathing frequency ( $BF_{peak}$ ) were calculated as the average value over the highest 30 seconds during the GWPT and the SRiT (prior to termination of the test).  $HR_{peak}$  was defined as the highest value during the tests.  $VO_{2peak}$  was normalized (ml/kg/min) by dividing absolute  $VO_{2peak}$  by body mass. The ventilator anaerobic threshold (VAT) was determined by the ventilatory equivalents method. When results were uncertain, the V-slope method was used to verify the VAT.<sup>19-22</sup>

### Field-based skill-related fitness testing.

Three field-based skill-related fitness tests were used in order to estimate the construct validity of the outcome measure shuttles from the SRiT.

Anaerobic performance was measured by the Muscle Power Sprint Test (MPST),<sup>14</sup> a valid ( $r > .72$ ; arm-cranking Wingate Anaerobic test) and reliable field test (ICCs  $> .95$ ) in youth with spina

bifida.<sup>14</sup> Participants had to sprint 15 m, marked by 2 lines, a total of 4 times. They had 10 seconds to turn and prepare for the next sprint, between every sprint. The time of a 15-m sprint was manually recorded with a stopwatch to one hundredth of a second. Power output for each of the 4 sprints was calculated using the following formula:

$$\text{Power} = (\text{total mass} \times \text{distance}^2) / \text{time}^{3,23}, \text{ where total mass included mass of the wheelchair and the child.}$$

Mean power was defined as the average power of the 4 sprints and was used as the outcome measure.

Agility was measured by the 10 × 5 Meter Sprint Test (10 × 5MST) and slalom test.<sup>14</sup> Both tests are valid ( $r = .93$ ) and reliable (ICC > .95) field tests in youth with spina bifida.<sup>14</sup> For the 10 × 5MST, participants had to sprint and turn 10 times continuously between 2 lines that were 5 m apart as fast as possible. During the slalom test, participants had to slalom between 4 cones that were 1.5 m apart, turn at the end, and sprint back as fast as possible and repeat the same procedure. Time taken to perform the tests was manually recorded with a stopwatch to one hundredth of a second and was used as the outcome measure.

## Data Analysis

**Power analysis.** We estimated the range of sample values given a hypothesized population ICC and sample size by a general method prior to data collection.<sup>24</sup> Assuming an ICC of .85, a sample size of 21 was needed for the lower limit of a 1-sided 95% CI to exceed a value of .70.<sup>25</sup>

**Criterion validity between the SRiT and the GWPT.** Criterion validity between the SRiT and the GWPT was evaluated by analyzing the cardiopulmonary responses (absolute and relative  $\text{VO}_{2\text{peak}}$ ,  $\text{VCO}_{2\text{peak}}$ ,  $\text{HR}_{\text{peak}}$ ,  $\text{VE}_{\text{peak}}$ ,  $\text{RER}_{\text{peak}}$ ,  $\text{VT}_{\text{peak}}$ ,  $\text{BF}_{\text{peak}}$ , and  $\text{VO}_2$  at VT) using 2-tailed  $t$  tests and Pearson correlation coefficients. First, normality of the data was checked with histograms and Q-Q plots. If there was uncertainty about the

normality of the data, bootstrapping with a bias-corrected accelerated CI was used to confirm the results. Also, linearity of relationships between the cardiopulmonary responses of the SRiT and the GWPT was assessed with scatterplots. Correlations ( $r$ ) were defined as weak if the lower bound of the 95% CI was less than .5, moderate if the lower bound of the 95% CI was .5 to .7, high if the lower bound of the 95% CI was .7 to .9, and excellent if the lower bound of the 95% CI was .9 to 1.0.<sup>26</sup>

**Construct validity outcome measure of shuttles achieved on the SRiT.** For construct validity, a stepwise regression was used to predict  $\text{VO}_{2\text{peak}}$  based on the number of shuttles achieved on the SRiT. Possible independent variables were personal factors (sex, age, weight, height, BMI), wheelchair features (wheelchair mass, tire pressure), and factors obtained during the SRiT ( $\text{HR}_{\text{peak}}$ , shuttles). First, linearity of relationships between  $\text{VO}_{2\text{peak}}$  and the independent variables was assessed with scatterplots and quantified with Pearson correlation coefficients. Subsequently, a weighted stepwise forward multiple regression analysis was performed to identify the independent variables that contributed to the prediction of  $\text{VO}_{2\text{peak}}$  during the SRiT. Variables were included with a  $P$  value less than .05 and excluded with a  $P$  value greater than .1. Multicollinearity was checked for by assuring a tolerance of greater than .1.<sup>27</sup> To present the accurateness of the regression on the individual level, we calculated prediction intervals for all participants.

The correlations between the number of achieved shuttles of the SRiT and the anaerobic performance (Mean Power of the MPST) and agility (seconds of the 10 × 5MST and seconds of the slalom test) were established by Pearson correlation coefficients. Additionally, possible relations between the number of achieved shuttles during the SRiT and personal factors (age, weight, height, BMI) and wheelchair features (wheelchair mass, tire pressure) were analyzed.

**Reliability of the SRiT.** Reliability was tested using the ICC model 2.1.A,<sup>28,29</sup> after checking for normality. Intraclass

correlation coefficients were defined as excellent if the lower bound of the 95% CI was greater than .80, high if the lower bound of the 95% CI was .7 to .8, and moderate if the lower bound of the 95% CI was .5 to .7.<sup>30</sup>

The measurement error for agreement was analyzed using the standard error of measurement for agreement ( $\text{SEM}_{\text{agreement}}$ ) and the smallest detectable change (SDC). The  $\text{SEM}_{\text{agreement}}$  was calculated by  $\text{SEM}_{\text{agreement}} = \sqrt{(\sigma_m^2 + \sigma_{\text{residual}}^2)}$ , in which  $\sigma_m^2$  accounts for the systematic errors between both measurements and  $\sigma_{\text{residual}}^2$  accounts for the random error.<sup>28,29</sup> The SDC was calculated by  $\text{SDC} = 1.96 \times \sqrt{2} \times \text{SEM}_{\text{agreement}}$ .<sup>29</sup>

In case of heteroscedasticity, checked by using a Bland-Altman plot,<sup>31</sup> the coefficient of variation was calculated as a measure of agreement.<sup>32,33</sup> The spreadsheet calculations by Hopkins were used with log transformation of the data.<sup>34</sup> Data were antilogged to obtain the 95% CI of the coefficient of variation.<sup>34</sup> A low coefficient of variation presents a more reliable measurement than a high coefficient of variation, with coefficients between 5% and 10% defined as acceptable.<sup>35-37</sup> For practical interpretation, we also created a Bland and Altman plot where we used the log transformed data for creating the limits of agreement.<sup>33,38</sup> After antilogging the data, the limits of agreement were placed in the Bland and Altman plot. These limits of agreement give an indication of the absolute agreement between the 2 measurements and can be interpreted as a true change, comparable to the SDC.<sup>33</sup>

## Results

A total of 51 children (30 boys/21 girls) were recruited, with a mean age of 13.6 years (SD = 3.7 y). For the validity part, 33 participants successfully completed both the GWPT and the SRiT. For the reliability part, 28 participants completed both the SRiT and the retest. Participants' characteristics, including level of lesion (classified according to the American Spinal Injury Association guidelines<sup>39</sup>) and modified Hoffer classification<sup>40</sup> are depicted in Table 1. A small number of participants were community or household ambulatory

## Wheelchair Shuttle Test for Aerobic Fitness in Youth With Spina Bifida

**Table 1.**  
Participant Characteristics

Characteristic	All Participants (N = 51)	Participants in Validity Testing (n = 33)	Participants in Reliability Testing (n = 28)
Age, year and month, mean (SD)	13.6 (3.7)	14.5 (3.1)	14.7 (3.3)
Body mass, kg, mean (SD)	48.8 (18.8)	53.5 (16.3)	51.3 (15.2)
Height, cm, mean (SD)	154.7 (21.3)	158.7 (16.7)	157.7 (17.2)
Body mass index, kg/m <sup>2</sup> , mean (SD)	22.9 (6.5)	24.6 (7.1)	24.2 (7.4)
Weight of wheelchair, kg, mean (SD)	19.2 (6.2)	18.3 (5.5)	20.0 (5.9)
Sex (no. of boys/no. of girls)	30/21	16/17	16/12
Type (no. open/no. closed)	47/4	32/1	27/1
Level of lesion <sup>39</sup> (no. of participants)			
Thoracic	11	5	7
Lumbar	39	28	21
Sacral	1	0	0
Ambulation level <sup>40</sup>			
Community ambulator	5	1	0
Household ambulator	6	4	4
Therapeutic ambulator	4	4	1
Nonambulator	36	24	23

and thus self-propelling a wheelchair for long distances or sports participation. The majority were therapeutic or nonambulatory, meaning that they were self-propelling a wheelchair during daily life. eTable 1, available at <https://academic.oup.com/ptj>, represents an overview of the data collection for the GWPT, SRiT and retest SRiT, including information regarding missing data. No adverse events occurred during testing.

### Criterion Validity of the SRiT and the GWPT

A total of 11 participants (mean age = 11.3 y; SD = 4.1; 7 boys and 4 girls) were excluded because they did not achieve maximal effort on the GWPT (n = 2) or the SRiT (n = 2) or on both (n = 7). Two participants for which the GWPT was too heavy and the SRiT was too difficult were 5 years and 9 months old and 6 years old, respectively.

No significant differences were found for HR<sub>peak</sub>, absolute VO<sub>2peak</sub>, relative VO<sub>2peak</sub>, VCO<sub>2peak</sub> and VE<sub>peak</sub> between the GWPT and SRiT. TV<sub>peak</sub> was significantly higher during the GWPT, while BF<sub>peak</sub> was significantly higher during the SRiT. High correlations between the

GWPT and the SRiT were found for absolute VO<sub>2peak</sub>, relative VO<sub>2peak</sub>, VCO<sub>2peak</sub>, VE<sub>peak</sub>, TV<sub>peak</sub>, and BF<sub>peak</sub> (lower bound of 95% CI:  $r > .7$ ), while weak correlations were found for HR<sub>peak</sub>, RER<sub>peak</sub>, and VO<sub>2</sub> at VT (lower bound of 95% CI:  $r = .2-.4$ ) (Tab. 2).

### Construct Validity

A total of 38 participants achieved the subjective criteria for maximal effort during the first SRiT and data of these participants were used for predicting absolute VO<sub>2peak</sub>.<sup>17</sup> Of the variables that concerned individual features (age, weight, height, BMI), wheelchair-user features (wheelchair mass, tire pressure), and outcomes obtained during the SRiT (shuttles, HR<sub>peak</sub>), only age ( $r = .47$ ), weight ( $r = .54$ ), height ( $r = .72$ ), and shuttles ( $r = .58$ ) correlated significantly with absolute VO<sub>2peak</sub>. Subsequently, height, weight, and shuttles were used as independent variables in the weighted multiple linear regression analysis. We used a weighted regression, because of heteroscedasticity during a normal multiple linear regression. The unstandardized predicted values were divided into 3 groups, using “1/variance” as the weighing factor. It

was possible to predict VO<sub>2peak</sub> for 77% by height, shuttles, and weight. The standard errors of estimates were 0.12, 0.14, and 0.26, respectively. Looking at individual prediction intervals, we saw a mean range of 0.734 L/min, with the smallest range being 0.51 L/min and the largest range being 1.10 L/min. The results of the weighted regression are presented in eTable 2, available at <https://academic.oup.com/ptj>.

Regarding the construct of the outcome measure shuttles, high correlations were found with the 10 × 5MST and the slalom test, both measuring agility (lower bound of 95% CI:  $r \geq -.8$ ). The MPST, measuring anaerobic power, correlated moderately (lower bound of 95% CI:  $r = .5$ ) with the number of shuttles. Looking at the personal variables, both height and BMI showed weak correlations with the number of shuttles (lower bound of 95% CI:  $r = .2$  and  $-.2$ ). Finally, wheelchair mass (lower bound of 95% CI:  $r = -.2$ ) correlated weakly with the number of shuttles as well (Tab. 3).

### Reliability

The reliability of the SRiT was high, with excellent ICCs for absolute VO<sub>2peak</sub>,

**Table 2.**

Results of the Graded Wheelchair Propulsion Test (GWPT) and the Shuttle Ride Test (SRiT)

Measures	GWPT, Mean (SD)	SRiT, Mean (SD)	Mean Difference (95% CI)	Correlation Coefficient (95% CI)
Absolute VO <sub>2peak</sub> (L/min) (n = 29 <sup>b</sup> )	1.1401 (0.400)	1.179 (0.302)	-0.039 (-0.121 to 0.042)	r = .848 (.699 to 0.926)
Relative VO <sub>2peak</sub> (mL/kg/min) (n = 29 <sup>b</sup> )	22.7 (6.9)	23.6 (6.5)	-0.9 (-2.4 to 0.5)	r = .841 (.686 to .923)
VCO <sub>2peak</sub> (L/min) (n = 29 <sup>b</sup> )	1.399 (0.565)	1.467 (0.513)	-0.068 (-0.153 to 0.017)	r = .918 (.832 to .961)
HR <sub>peak</sub> (bpm) (n = 33)	185 (16)	188 (14)	-3 (-8 to 1)	r = .683 (.423 to .839)
VE <sub>peak</sub> (22 L/min) (n = 29 <sup>b</sup> )	53.21 (22.13)	54.31 (18.15)	-1.09 (-4.64 to 2.46)	r = .911 (.818 to .957)
RER <sub>peak</sub> (n = 29 <sup>b</sup> )	1.22 (0.17)	1.25 (0.20)	-0.03 (-0.09 to 0.04)	r = .592 (.288 to .787)
VT <sub>peak</sub> (L) (n = 29 <sup>b</sup> )	1.012 (0.346)	0.965 (0.348)	0.047 (0.013 to 0.081)	r = .967 (.931 to .984)
BF <sub>peak</sub> (n = 28 <sup>c</sup> )	56.50 (16.14)	61.15 (13.25)	-4.65 (-7.37 to -1.93)	r = .904 (.802 to .954)
VO <sub>2</sub> at VT (L/min) (n = 28 <sup>c</sup> )	0.757 (0.290)	0.730 (0.262)	0.027 (-0.075 to 0.128)	r = .552 (.226 to .767)

<sup>a</sup>bpm = beats per minute, BF<sub>peak</sub> = peak breathing frequency, HR<sub>peak</sub> = peak heart rate, RER<sub>peak</sub> = peak respiratory exchange ratio, VCO<sub>2peak</sub> = peak carbon dioxide production, VE<sub>peak</sub> = peak minute ventilation, VO<sub>2</sub> at VT = oxygen uptake at ventilatory threshold, VO<sub>2peak</sub> = peak oxygen uptake, VT<sub>peak</sub> = peak tidal volume.

<sup>b</sup>Missing data for respiratory gas analysis measurements in 4 participants.

<sup>c</sup>Missing data for respiratory gas analysis measurements in 5 participants.

VCO<sub>2peak</sub>, TV<sub>peak</sub>, VE<sub>peak</sub>, and the number of achieved shuttles (lower bound of 95% CI: ICC > .9), high ICCs for BF<sub>peak</sub> and relative VO<sub>2peak</sub> (lower bound of 95% CI: ICC = .8), and moderate ICCs for HR<sub>peak</sub>, RER<sub>peak</sub>, and VO<sub>2</sub> at VT (lower bound of 95% CI: ICC = .5–.6).

For agreement, the SDC for the number of achieved shuttles is 1.5. The coefficient of variation for absolute and relative VO<sub>2peak</sub> is 6.2% (95% CI = 5.0–8.2) and 6.4% (95% CI = 5.2–8.6), respectively. The limits of agreement are for both absolute and relative VO<sub>2peak</sub> ± 0.20 x the mean of the first and second measurements (Tab. 4; Figs. 1 and 2).

### Discussion

This study focused on the validity and reliability of the SRiT in youth with spina bifida who use a wheelchair for mobility and sport. To our knowledge, only one other study so far has investigated the validity of the SRiT in youth who use wheelchairs. Verschuren et al<sup>8</sup> compared the SRiT with the graded arm exercise test (GAET) in youth with CP. They found a significantly higher HR<sub>peak</sub> and VE<sub>peak</sub> during the SRiT and hence questioned the GAET as a gold standard to measure cardiorespiratory demands in children who use wheelchairs.<sup>8</sup> We therefore considered the wheelchair propulsion test (GWPT) as our gold

standard to ensure specificity of testing, as we found a higher HR<sub>peak</sub> and VO<sub>2peak</sub> during the GWPT compared to the GAET in an earlier study.<sup>7</sup> The differences we found in TV<sub>peak</sub> and BF<sub>peak</sub> between the SRiT and GWPT may be explained by the differences in test performance. During the GWPT, the child has to propel with a continuous speed and increasing load, while during the SRiT, the child has to increase his or her speed. Our hypothesis is that this difference in test performance may affect the breathing pattern and thus explaining the increase of BF during the SRiT and increase of TV during the GWPT. Future research may clarify these different physiologic responses during incremental exercise testing protocols in youth who use wheelchairs.

Regarding cardiopulmonary responses of youth with spina bifida, we observed higher HR<sub>peak</sub>, RER<sub>peak</sub>, and VE<sub>peak</sub> compared to youth with CP, with similar VO<sub>2peak</sub> results.<sup>8</sup> When comparing our cardiopulmonary responses with youth with osteogenesis imperfecta, we observed lower HR<sub>peak</sub>, RER<sub>peak</sub>, and higher VE<sub>peak</sub>, again with comparable VO<sub>2peak</sub> values.<sup>9</sup> It would be interesting to further analyze these cardiopulmonary responses in youth with different diagnosis, to gain a deeper insight in exercise physiology of youth who use a wheelchair for mobility and sport, so

these findings can be interpreted more adequately. Moreover, this would also help to understand which objective criteria for maximal aerobic exercise testing should be used in this population. Until now, the applicability of the existing objective criteria for maximal aerobic exercise testing (HR<sub>peak</sub> ≥ 180/min and RER<sub>peak</sub> > 0.99)<sup>17</sup> is unclear for youth who use a wheelchair for mobility and sport because of the smaller muscle mass in the arms compared to the legs. In this study we used subjective criteria for maximal exercise testing<sup>17</sup> to conclude if a child performed maximal at either the GWPT or the SRiT. We only included data in the analyses, if the subjective criteria were met. There were no specific characteristics regarding participants who did not meet the subjective criteria, so unfortunately we were not able to conclude in which children the SRiT cannot be used for maximal cardiorespiratory exercise testing. We also tried to use the OMNI scale of perceived exertion<sup>41</sup>; unfortunately, these results were unreliable due to the cognitive impairments often present in youth with spina bifida.

To our knowledge, no study in youth who use a wheelchair for mobility and sport tried to predict VO<sub>2peak</sub> using the number of achieved shuttles during the SRiT so far. A recent meta-analysis concerning the original 20-m Shuttle

## Wheelchair Shuttle Test for Aerobic Fitness in Youth With Spina Bifida

**Table 3.**

Correlations of Number of Achieved Shuttles During the Shuttle Ride Test With Skill-Related Fitness Tests, Personal Factors, and Wheelchair Features<sup>a</sup>

	Correlation With:								
	MPST (95% CI)	10 × 5MST (95% CI)	Slalom Test (95% CI)	Age (95% CI)	Height (95% CI)	Weight (95% CI)	BMI (95% CI)	Wheelchair Mass (95% CI)	Tire Pressure (95% CI)
Shuttles	.73 (.54 to .85)	-.92 (-.96 to -.86)	-.89 (-.94 to -.80)	.11 (-.20 to .41)	.45 (.16 to .66)	-.12 (-.19 to .41)	-.47 (-.70 to -.19)	-.54 (-.72 to -.18)	-.15 (-.44 to .26)

<sup>a</sup>BMI = body mass index, MPST = Muscle Power Sprint Test, 10 × 5MST = 10 × 5 Meter Sprint Test.

**Table 4.**

Outcome Reliability Data for Shuttle Ride Test (SRIT)<sup>a</sup>

Measure	SRIT 1 Mean (SD)	SRIT 2 Mean (SD)	ICC 2.1.A (95% CI)	SEM <sub>agreement</sub>	Smallest Detectable Change	Coefficient of Variation (95% CI)
HR <sub>peak</sub> (bpm) (n = 27 <sup>b</sup> )	189 (13)	188 (15)	.77 (.56-.90)	7	18	
VCO <sub>2peak</sub> (L/min) (n = 25 <sup>c</sup> )	1.436 (0.482)	1.505 (0.549)	.97 (.94-.99)	0.083	0.232	
RER <sub>peak</sub> (n = 25 <sup>c</sup> )	1.22 (0.13)	1.21 (0.18)	.78 (.52-.91)	0.06	0.18	
BF <sub>peak</sub> (n = 26 <sup>d</sup> )	59.50 (14.27)	57.82 (14.90)	.92 (.83-.96)	4.195	11.629	
TV <sub>peak</sub> (L) (n = 27 <sup>b</sup> )	0.981 (0.333)	1.024 (0.364)	.95 (.89-.98)	0.083	0.232	
VE <sub>peak</sub> (L/min) (n = 25 <sup>c</sup> )	53.99 (18.07)	56.95 (21.42)	.96 (.91-.98)	9.314	25.818	
VO <sub>2</sub> at VT (L/min) (n = 24 <sup>e</sup> )	0.714 (0.257)	0.742 (0.273)	.773 (.55-.90)	0.126	0.351	
Shuttles (n = 28)	14 (4)	14 (4)	.98 (.96-.99)	0.5	1.5	
VO <sub>2peak</sub> (L/min) (n = 25 <sup>c</sup> )	1.193 (0.362)	1.245 (0.387)	.96 (.92-.98)			6.2% (5.0-8.2)
VO <sub>2peak</sub> (mL/kg/min) (n = 25 <sup>c</sup> )	24.1 (6.0)	25.1 (6.8)	.93 (.84-.97)			6.4% (5.2-8.6)

<sup>a</sup>bpm = beats per minute, BF<sub>peak</sub> = peak breathing frequency, HR<sub>peak</sub> = peak heart rate, ICC 2.1.A = intraclass correlation coefficient model 2.1.A, RER<sub>peak</sub> = peak respiratory exchange ratio, SEM<sub>agreement</sub> = standard error of measurement for agreement, TV<sub>peak</sub> = peak tidal volume, VCO<sub>2peak</sub> = peak carbon dioxide production, VE<sub>peak</sub> = peak minute ventilation, VO<sub>2</sub> at VT = oxygen uptake at ventilatory threshold, VO<sub>2peak</sub> = peak oxygen uptake.

<sup>b</sup>Missing data for respiratory gas analysis measurements in 1 participant.

<sup>c</sup>Missing data for respiratory gas analysis measurements in 3 participants.

<sup>d</sup>Missing data for respiratory gas analysis measurements in 2 participants.

<sup>e</sup>Missing data for respiratory gas analysis measurements in 4 participants.

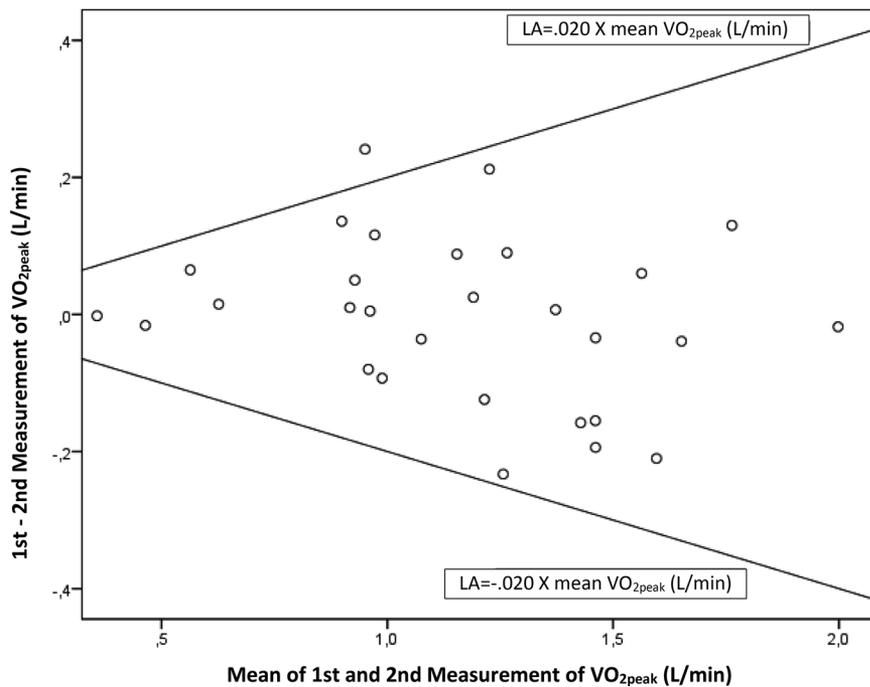
Run Test for typically developing children showed a moderate to high criterion-related validity for estimating VO<sub>2peak</sub>.<sup>42</sup> However, Castro-Pinero et al<sup>43</sup> stated that existing equations for estimating VO<sub>2peak</sub> should not be used at an individual level in typically developing children. A study in adults was only able to predict VO<sub>2peak</sub> for 59% of the variance using the number of exercise stages during an incremental SRIT<sup>44</sup> and another study in adults concluded strong reservations about predicting VO<sub>2peak</sub> in adults who use wheelchairs.<sup>12</sup> Even though we were able to explain 77% of the variance, relatively large prediction intervals were present, indicating large errors when using the prediction equation at

an individual level. Of course, our relatively small sample size of 38 should be taken into account, so our results should be interpreted as tentative.<sup>45</sup> This is why, for now, we recommend *not* to use this prediction equation and advise to use a mobile gas analysis system to measure VO<sub>2peak</sub> in children with spina bifida when interested in aerobic fitness.

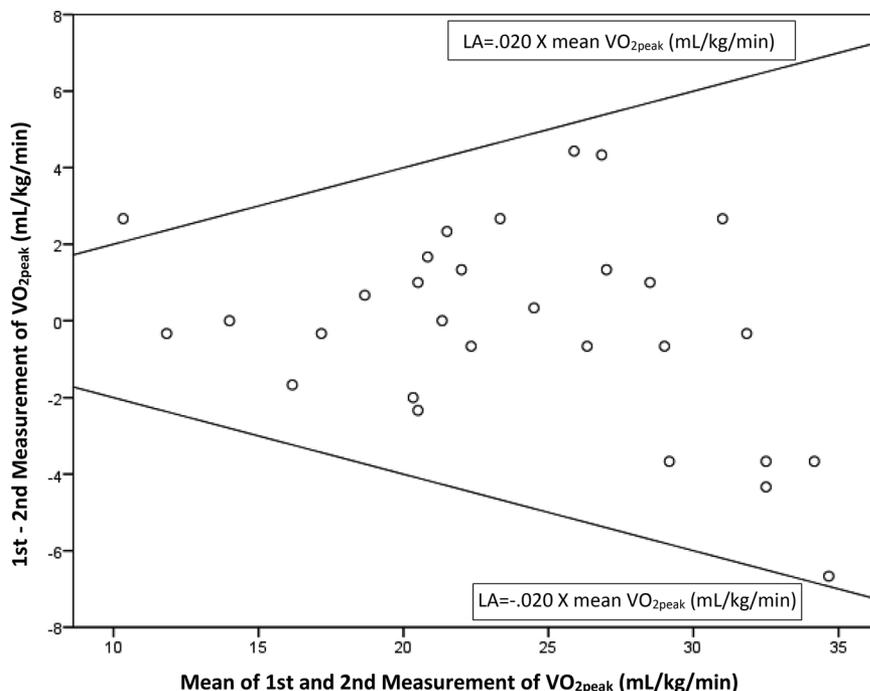
We then tried to clarify the construct of the outcome measure shuttle, mostly used by pediatric physical therapists because they do not have the availability of a mobile gas analysis system. Unfortunately, it was not possible to explain which independent variables contribute to the number of

achieved shuttles using a multiple linear regression, due to multicollinearity between the skill-related fitness tests. The moderate to high correlations between the number of achieved shuttles and both anaerobic performance and agility confirm the hypothesis generated by Verschuren et al<sup>8</sup> and Vanlandewijck et al<sup>12</sup> about the importance of anaerobic performance and agility during the SRIT. It underlines the importance of mastering wheelchair skills as deceleration, turning and acceleration, next to optimizing VO<sub>2peak</sub>. Another interesting subject would be to analyze whether the increase in VO<sub>2</sub> is equal during every incremental shuttle of the SRIT. This might help to analyze which part of the SRIT is explained by VO<sub>2</sub> uptake and

## Wheelchair Shuttle Test for Aerobic Fitness in Youth With Spina Bifida



**Figure 1.** Limits of agreement (LA) for absolute peak oxygen uptake ( $VO_{2peak}$ ).



**Figure 2.** Limits of agreement (LA) for relative peak oxygen uptake ( $VO_{2peak}$ ).

which part may be explained by, for example, anaerobic performance or agility.

It was interesting to see the significant negative correlation between wheelchair

mass and number of achieved shuttles, indicating the importance of light-weight wheelchairs. In our previous study about skill-related fitness tests in youth with spina bifida,<sup>14</sup> we also found

that a lighter wheelchair mass contributed to the distance traveled during 1 push. Literature about adults confirm the relevance of light weight wheelchairs.<sup>46</sup> Interestingly, some participants had relatively heavy wheelchairs, so wheelchair mass should be considered more carefully when providing a wheelchair to a child. Unfortunately, we were not able to take wheelchair features as rolling resistance, internal resistance and the wheelchair configuration into account.<sup>47-49</sup> These aspects are difficult to measure in this population, as all children have individually adjusted wheelchairs. Future research may take these wheelchair features into account.

The results regarding the reliability of the achieved shuttles during the SRiT are comparable to those for youth who have CP and osteogenesis imperfecta and use wheelchairs.<sup>8,50</sup> The SDC for the number of achieved shuttles (SDC = 1.5) is similar to that of youth who have CP and use wheelchairs (SDC = 1.4) and slightly better than that of youth who have osteogenesis imperfecta and use wheelchairs (SDC = 1.9).<sup>8,50</sup>

In conclusion, the SRiT is a highly valid field test for measuring  $VO_{2peak}$  in youth with spina bifida who use a wheelchair for mobility and sport, when applying a mobile gas analysis system during testing. If a mobile gas analysis system is available, both the SRiT and the GWPT can be used to measure  $VO_{2peak}$ . It is *not* possible to predict  $VO_{2peak}$  accurately using the number of achieved shuttles. For pediatric physical therapists using the metric shuttles, the number of achieved shuttles represents aerobic fitness and is moderately correlated with anaerobic performance and highly correlated with agility. Because the SRiT is highly reliable and has a good SDC for the number of achieved shuttles, the SRiT can be used to monitor effectiveness of interventions to improve aerobic performance in youth with spina bifida who use a wheelchair for mobility and sport.

### Author Contributions and Acknowledgements

Concept/idea/research design: M. Bloemen, J.F. de Groot, T. Takken  
Writing: M. Bloemen, J.F. de Groot, T. Takken

Data collection: M. Bloemen, J. Benner  
 Data analysis: M. Bloemen, J.F. de Groot, J. Benner, C. Kruitwagen, T. Takken  
 Project management: M. Bloemen, J.F. de Groot, J. Benner, C. Kruitwagen, T. Takken  
 Fund procurement: J.F. de Groot, T. Takken  
 Providing participants: M. Bloemen,  
 Providing facilities/equipment: T. Takken  
 Providing institutional liaisons: T. Takken  
 Consultation (including review of manuscript before submitting): J.F. de Groot, F.J.G. Backx, T. Takken

The authors thank all of the children, adolescents, and parents for their enthusiastic participation. They also thank all of the students who participated in this study as part of their theses.

### Ethics Approval

The study procedures were approved by the Medical Ethics Committee of the University Medical Center Utrecht, Utrecht, the Netherlands. Children 12 years old and older and all parents were required by Dutch law and regulations to sign informed consent forms.

### Funding

This study was funded by the Foundation Innovation Alliance-Regional Attention and Action for Knowledge Circulation (SIA RAAK) (project no. 2011-13-35P) and by a personal doctoral grant (M. Bloemen) from HU University of Applied Sciences Utrecht.

### Disclosures/Presentations

The authors completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest and reported no conflicts of interest.

DOI: 10.1093/ptj/pzx075

### References

- Bouchard C, Shephard RJ, Stephens T. Physical activity, fitness and health: the model and key concepts. In: *Physical Activity, Fitness, and Health Consensus Statement*. Champaign, IL: Human Kinetics; 1994.
- Rowland JL, Fragala-Pinkham M, Miles C, O'Neil ME. The scope of pediatric physical therapy practice in health promotion and fitness for youth with disabilities. *Pediatr Phys Ther*. 2015;27(1):2-15.
- Bongers BC, van Brussel M, Hulzebos EHJ, Takken T. *Pediatric Norms for Cardiopulmonary Exercise Testing in Relation to Sex and Age*. 2nd ed. 's-Hertogenbosch: BOXPress; 2014.
- Schoenmakers MA, de Groot JF, Gorter JW, Hillaert JL, Helders PJ, Takken T. Muscle strength, aerobic capacity and physical activity in independent ambulating children with lumbosacral spina bifida. *Disabil Rehabil*. 2009;31(4):259-266.
- De Groot JF, Takken T, Schoenmakers MA, Vanhees L, Helders PJ. Limiting factors in peak oxygen uptake and the relationship with functional ambulation in ambulating children with spina bifida. *Eur J Appl Physiol*. 2008;104(4):657-665.
- Buffart LM, Roebroek ME, Rol M, Stam HJ, van den Berg-Emons RJ, Transition Research Group South-West Netherlands. Triad of physical activity, aerobic fitness and obesity in adolescents and young adults with myelomeningocele. *J Rehabil Med*. 2008;40(1):70-75.
- Bloemen MA, de Groot JF, Backx FJ, Westerveld RA, Takken T. Arm cranking versus wheelchair propulsion for testing aerobic fitness in children with spina bifida who are wheelchair dependent. *J Rehabil Med*. 2015;47(5):432-437.
- Verschuren O, Zwinkels M, Ketelaar M, Reijnders-van Son F, Takken T. Reproducibility and validity of the 10-meter shuttle ride test in wheelchair-using children and adolescents with cerebral palsy. *Phys Ther*. 2013;93(7):967-974.
- Bongers BC, Rijks EB, Harsevoort AG, Takken T, van Brussel M. 10-m shuttle ride test in youth with osteogenesis imperfecta who use wheelchairs: Feasibility, reproducibility, and physiological responses. *Phys Ther*. 2016;96(5):679-686.
- Leger LA, Mercier D, Gadoury C, Lambert J. The multistage 20 metre shuttle run test for aerobic fitness. *J Sports Sci*. 1988;6(2):93-101.
- Goosey-Tolfrey VL, Tolfrey K. The multi-stage fitness test as a predictor of endurance fitness in wheelchair athletes. *J Sports Sci*. 2008;26(5):511-517.
- Vanlandewijck Y, van de Vliet P, Verellen J, Theisen D. Determinants of shuttle run performance in the prediction of peak VO<sub>2</sub> in wheelchair users. *Disabil Rehabil*. 2006;28(20):1259-1266.
- Goosey-Tolfrey VL, Leicht CA. Field-based physiological testing of wheelchair athletes. *Sports Med*. 2013;43(2):77-91.
- Bloemen MAT, Takken T, Backx FJG, Vos M, Kruitwagen CJLL, de Groot JF. Validity and reproducibility of skill-related fitness tests for wheelchair-using youth with spina bifida. *Arch Phys Med Rehabil*. 2017;98:1097-1103.
- Dosa NP, Foley JT, Eckrich M, Woodall-Ruff D, Liptak GS. Obesity across the lifespan among persons with spina bifida. *Disabil Rehabil*. 2009;31(11):914-920.
- Medbo JI, Mamen A, Welde B, von Heimburg E, Stokke R. Examination of the metamax I and II oxygen analysers during exercise studies in the laboratory. *Scand J Clin Lab Invest*. 2002;62(8):585-598.
- Rowland T. *Developmental Exercise Physiology*. Champaign, IL: Human Kinetics; 1996.
- Paap D, Takken T. Reference values for cardiopulmonary exercise testing in healthy adults: A systematic review. *Expert Rev Cardiovasc Ther*. 2014;12(12):1439-1453.
- Visschers NC, Hulzebos EH, van Brussel M, Takken T. Comparing four non-invasive methods to determine the ventilatory anaerobic threshold during cardiopulmonary exercise testing in children with congenital heart or lung disease. *Clin Physiol Funct Imaging*. 2015;35(6):451-459.
- American Thoracic Society, American College of Chest Physicians. ATS/ACCP statement on cardiopulmonary exercise testing. *Am J Respir Crit Care Med*. 2003;167(2):211-277.
- Ohuchi H, Nakajima T, Kawade M, Matsuda M, Kamiya T. Measurement and validity of the ventilatory threshold in patients with congenital heart disease. *Pediatr Cardiol*. 1996;17(1):7-14.
- Beaver WL, Wasserman K, Whipp BJ. A new method for detecting anaerobic threshold by gas exchange. *J Appl Physiol* (1985). 1986;60(6):2020-2027.
- Verschuren O, Takken T, Ketelaar M, Gorter JW, Helders PJ. Reliability for running tests for measuring agility and anaerobic muscle power in children and adolescents with cerebral palsy. *Pediatr Phys Ther*. 2007;19(2):108-115.
- Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420-428.
- Verschuren O, Takken T, Ketelaar M, Gorter JW, Helders PJ. Reliability and validity of data for 2 newly developed shuttle run tests in children with cerebral palsy. *Phys Ther*. 2006;86(8):1107-1117.
- Cohen J. *Statistical Power Analysis for the Behavioural Sciences, Second Edition*. Hillsdale, NJ, UK: Lawrence Erlbaum Associates; 1988.
- Field A. *Discovering Statistics using SPSS*. Second ed. London, United Kingdom: SAGA Publications Ltd; 2006.
- de Vet H, Terwee C, Mokkink L, Knol D. *Measurement in Medicine. Practical Guide to Biostatistics and Epidemiology*. New York, NY: Cambridge University Press; 2013.
- de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol*. 2006;59(10):1033-1039.
- Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Practice*. Norwalk, CT: Appleton and Lange; 2000.
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1(8476):307-310.

- 32 Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med.* 1998;26(4):217-238.
- 33 Stekelenburg CM, van der Wal MB, Knol DL, de Vet HC, van Zuijlen PP. Three-dimensional digital stereophotogrammetry: A reliable and valid technique for measuring scar surface area. *Plast Reconstr Surg.* 2013;132(1):204-211.
- 34 Hopkins WG. A new view of statistics. A new view of statistics Web site. <http://sports.org/resource/stats/>. Published 2000. Updated 2013. Accessed February, 2016.
- 35 Figueroa-Colon R, Hunter GR, Mayo MS, Aldridge RA, Goran MI, Weinsier RL. Reliability of treadmill measures and criteria to determine VO<sub>2</sub>max in prepubertal girls. *Med Sci Sports Exerc.* 2000;32(4):865-869.
- 36 Johnston KN, Jenkins SC, Stick SM. Repeatability of peak oxygen uptake in children who are healthy. *Pediatr Phys Ther.* 2005;17(1):11-17.
- 37 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 training? *Phys Ther.* 2011;91(2):267-276.
- 38 Euser AM, Dekker FW, le Cessie S. A practical approach to bland-altman plots and variation coefficients for log transformed variables. *J Clin Epidemiol.* 2008;61(10):978-982.
- 39 Maynard FM, Jr, Bracken MB, Creasey G, et al. International standards for neurological and functional classification of spinal cord injury. american spinal injury association. *Spinal Cord.* 1997;35(5):266-274.
- 40 Schoenmakers MA, Gulmans VA, Gooskens RH, Helders PJ. Spina bifida at the sacral level: More than minor gait disturbances. *Clin Rehabil.* 2004;18(2):178-185.
- 41 Robertson RJ, Goss FL, Boer NF, et al. Children's OMNI scale of perceived exertion: Mixed gender and race validation. *Med Sci Sports Exerc.* 2000;32(2):452-458.
- 42 Mayorga-Vega D, Aguilar-Soto P, Viciana J. Criterion-related validity of the 20-M shuttle run test for estimating cardiorespiratory fitness: A meta-analysis. *J Sports Sci Med.* 2015;14(3):536-547.
- 43 Castro-Pinero J, Artero EG, Espana-Romero V, et al. Criterion-related validity of field-based fitness tests in youth: A systematic review. *Br J Sports Med.* 2010;44(13):934-943.
- 44 Vanderthommen M, Francaux M, Colinet C, et al. A multistage field test of wheelchair users for evaluation of fitness and prediction of peak oxygen consumption. *J Rehabil Res Dev.* 2002;39(6):685-692.
- 45 Babyak MA. What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models. *Psychosom Med.* 2004;66(3):411-421.
- 46 Cooper RA, Koontz AM, Ding D, Kelleher A, Rice I, Cooper R. Manual wheeled mobility—current and future developments from the human engineering research laboratories. *Disabil Rehabil.* 2010;32(26):2210-2221.
- 47 Mason BS, van der Woude LH, Goosey-Tolfrey VL. The ergonomics of wheelchair configuration for optimal performance in the wheelchair court sports. *Sports Med.* 2013;43(1):23-38.
- 48 van der Woude LH, Veeger HE, Dallmeijer AJ, Janssen TW, Rozendaal LA. Biomechanics and physiology in active manual wheelchair propulsion. *Med Eng Phys.* 2001;23(10):713-733.
- 49 van der Woude LH, de Groot S, Janssen TW. Manual wheelchairs: Research and innovation in rehabilitation, sports, daily life and health. *Med Eng Phys.* 2006;28(9):905-915.
- 50 Bongers BC, Rijks EB, Harsevoort AG, Takken T, van Brussel M. 10-m shuttle ride test in youth with osteogenesis imperfecta who use wheelchairs: Feasibility, reproducibility, and physiological responses. *Phys Ther.* 2015.