The Quality Function Measure: reliability and discriminant validity of a new measure of quality of gross motor movement in ambulatory children with cerebral palsy

F VIRGINIA WRIGHT[1,2] | PETER ROSENBAUM[3,4] | DARCY FEHLINGS[5,6] | RONIT MESTERMAN[7] | UTE BREUER[8] | MARIE KIM[9]

1 Bloorview Research Institute, Toronto, ON; **2** Department of Physical Therapy, University of Toronto, Toronto, ON; **3** *CanChild* Centre for Disability Research, Hamilton, ON; **4** Department of Pediatrics, McMaster University, Hamilton, ON; **5** Holland Bloorview Kids Rehabilitation Hospital, Toronto, ON; **6** Department of Pediatrics, University of Toronto, Toronto, ON; **7** Developmental Pediatric Rehabilitation and Autism Spectrum Disorder Services, McMaster University, Hamilton, ON, Canada. **8** Dr von Haunersches Children's Hospital, Munich, Germany. **9** ErinoakKids Centre for Treatment and Development, Mississauga, ON, Canada.

Correspondence to F Virginia Wright at Bloorview Research Institute, 150 Kilgour Rd, Toronto, ON M4G 1R8, Canada. E-mail: vwright@hollandbloorview.ca

PUBLICATION DATA
Accepted for publication 14th January 2014.
Published online

ABBREVIATIONS

GMPM	Gross Motor Performance Measure
LoA	Limits of agreement
MDC	Minimal detectable change
QFM	Quality Function Measure

AIM Optimizing movement quality is a common rehabilitation goal for children with cerebral palsy (CP). The new Quality Function Measure (QFM) – a revision of the Gross Motor Performance Measure (GMPM) – evaluates five attributes: Alignment, Co-ordination, Dissociated movement, Stability, and Weight-shift, for the Gross Motor Function Measure (GMFM) Stand and Walk/Run/Jump items. This study evaluated the reliability and discriminant validity of the QFM.

METHOD Thirty-three children with CP (17 females, 16 males; mean age 8y 11mo, SD 3y 1mo; Gross Motor Function Classification System [GMFCS] levels I [*n*=17], II [*n*=7], III [*n*=9]) participated in reliability testing. Each did a GMFM Stand/Walk assessment, repeated 2 weeks later. Both GMFM assessments were videotaped. A physiotherapist assessor pair independently scored the QFM from an assigned child's GMFM video. GMFM data from 112 children. That is, (GMFCS I [*n*=38], II [*n*=27], III [*n*=47]) were used for discriminant validity evaluation.

RESULTS QFM mean scores varied from 45.0% (SD 27.2; Stability) to 56.2% (SD 27.5; Alignment). Reliability was excellent across all attributes: intraclass correlation coefficients (ICCs) ≥0.97 (95% confidence intervals [CI] 0.95–0.99), interrater ICCs ≥0.89 (95% CI 0.80–0.98), and test–retest ICCs ≥0.90 (95% CI 0.79–0.99). QFM discriminated qualitative attributes of motor function among GMFCS levels (maximum $p<0.05$).

INTERPRETATION The QFM is reliable and valid, making it possible to assess how well young people with CP move and what areas of function to target to enhance quality of motor control.

Impaired gross motor development and function are defining features of cerebral palsy (CP).[1] Interventions aim to optimize what the child can do (functional skills) and how they do it (movement quality).[2] Both are important prerequisites for advanced motor skills related to activity and participation. Enhancement of quality of movement may be an important precursor to attainment of new gross motor skills.[2] While function is the ultimate goal, identifying challenges in quality of movement helps to guide therapists to develop individualized therapy approaches to enhance function. Without measuring movement, quality changes can only be inferred from improvements in other aspects of function, such as timed walk tests, acquisition of new motor skills, or indications of greater functional skill independence. Clinicians and researchers, therefore, remain uncertain about what aspects of quality are changing and when, how, and why. 'No difference' findings in trials that failed to assess quality of movement may have missed important changes in this underlying component of motor skill.

Instrumented systems quantify selected attributes of movement quality (e.g. force, amplitude, speed), but their use is limited to specialized centres. Paediatric physiotherapists typically rate quality of performance through clinical observation and Gestalt perceptions of movement.[3,4] The absence of appropriate clinical instruments to evaluate movement quality was the impetus to the development of the Gross Motor Performance Measure (GMPM),[3] created through an expert consensus approach as a companion to the original Gross Motor Function Measure (GMFM-88).[5] The GMPM was designed to quantify key aspects of movement quality and evaluate change in children with CP and acquired brain injury.[6] Five attributes (Alignment, Co-ordination, Dissociated movement, Stability, and

Weight-shift) were systematically identified by clinical experts to evaluate movement quality for 20 GMFM-88 items.[6] Three appropriate quality attributes were identified and linked to each item. The GMPM demonstrated excellent interrater and test–retest reliability (i.e. intraclass correlation coefficients [ICCs] >0.80) with in-person rating by GMPM-experienced raters.[7–9] It discriminated among children with 'mild', 'moderate', and 'severe' CP[10] and those judged to be stable versus those showing quality changes;[10] and it detected change with interventions (orthopaedic surgery, rhizotomy, and ankle orthoses).[11,12]

The GMPM remains the only published measure that addresses multiple components of quality of gross motor skills for children with CP. Other published upper or lower extremity/total body quality of movement scales are outlined in Table SI (online supporting information). Three concerns about the GMPM are its lack of item-specific response option descriptions, the need for highly experienced raters,[8] and the small representation of motor skills (20 GMFM-88 items). While the GMPM may be suitable for a therapist who is working with a child on basic motor skills that have quality components to them (e.g. Co-ordination and Dissociated movement are relevant attributes to assess for lying/rolling and crawling skills), its evaluation of movement quality across *all* aspects of motor function does not align well with goals of children who are working specifically on ambulation-based skills. For these children in Gross Motor Function Classification System (GMFCS) levels I to III, evaluation of movement quality should be intensely focused on standing and walking skills with a sufficient number of items to give a complete picture of their movement quality.

In response to the need for a well-developed measure of quality of motor function for ambulatory children with CP, two GMPM development team members (FVW and PR) led the creation of a new version of the GMPM. This new version was designed to focus on thorough assessment of movement quality across the 33 standing, walking, running, and jumping items in the GMFM-66 with children ages 4 and up in GMFCS levels I, II, and III.[13] We renamed this version the Quality Function Measure (QFM) to differentiate it from GMPM, and to emphasize that 'quality' (of gross motor skills) is being evaluated. In designing the QFM, rating movement quality 'with' and 'without' performance cues replaced the 'pathology'/'no pathology' distinction of the GMPM. This was done to measure the child's overall 'capacity' by integrating optimized (cued) quality with spontaneous (natural) 'performance'. As quality of movement refinements occur, one might expect to see natural performance styles becoming more like cued performance (i.e. closure of the performance gap), and item mean scores rise.

In a phase of work conducted by our team before this measurement study, the QFM development process began by reworking GMPM items and attributing allocations through consensus methods with 21 expert paediatric physiotherapists from Canada, the US, France, and Britain at an international CP meeting. The three quality attributes judged by these experts to be most appropriate (chosen from the five GMPM attributes) were assigned to each GMFM-66 Stand and Walk item. Next, item-specific response options were developed for the QFM to make rating easier and more consistent (Fig. 1). The format and wording of these new response options were tested with several local physiotherapists after which the revised options were reviewed by 10 physiotherapists from the original consensus group. Given the detail in these item-specific response options and the challenges observed when physiotherapists simultaneously scored the three GMPM attributes/items during in-person GMFM test sessions,[9,14,15] we decided that the QFM would be scored from a video recording of each child's GMFM-66 assessment.

This paper reports the results from the first QFM validation study, designed to evaluate intrarater, interrater, and test–retest reliability in children with CP, and to assess its discriminant validity in relation to GMFCS levels.

What this paper adds

- The Quality Function Measure (QFM) evaluates key quality attributes of motor function in ambulatory children with CP.
- The measure has excellent interrater, intrarater, and test–retest reliability.
- The QFM attributes differentiate children by GMFCS level.
- The QFM provides a 'higher-power' assessment of movement quality of young people with CP.

METHOD

We conducted a longitudinal measurement study with baseline and retest sessions. Local Research Ethics Board Committees at Holland Bloorview Kids Rehabilitation Hospital, McMaster Children's Hospital, and ErinoakKids (all in Ontario, Canada) approved the study. Written informed assent/consent was obtained from the children and parents.

Participants

A convenience sample was consecutively drawn from physiotherapist caseloads of children at the three participating treatment centres. A child/adolescent was eligible if they (1) were aged 4 to 18 years inclusive with CP in GMFCS levels I, II, or III (2) could follow instructions on the GMFM Stand and Walk/Run/Jump dimensions (as judged by their treating physiotherapist); (3) were at least 3 months post-botulinum toxin injections with no injections planned during the retest interval, and (4) had not had orthopaedic surgery within the last 6 months.

Additional children who met the first two eligibility criteria were entered into the discriminant validity component of the validation study to enhance sample size; these were primarily children in our QFM responsiveness study (i.e. undergoing botulinum toxin injections or orthopaedic surgery, or taking part in conductive education or in a barefoot/ankle-foot orthosis gait comparison). All QFM reliability study children's baseline scores were included in the discriminant validity analysis, as well as scores from four children who attended for a single GMFM/QFM assessment.

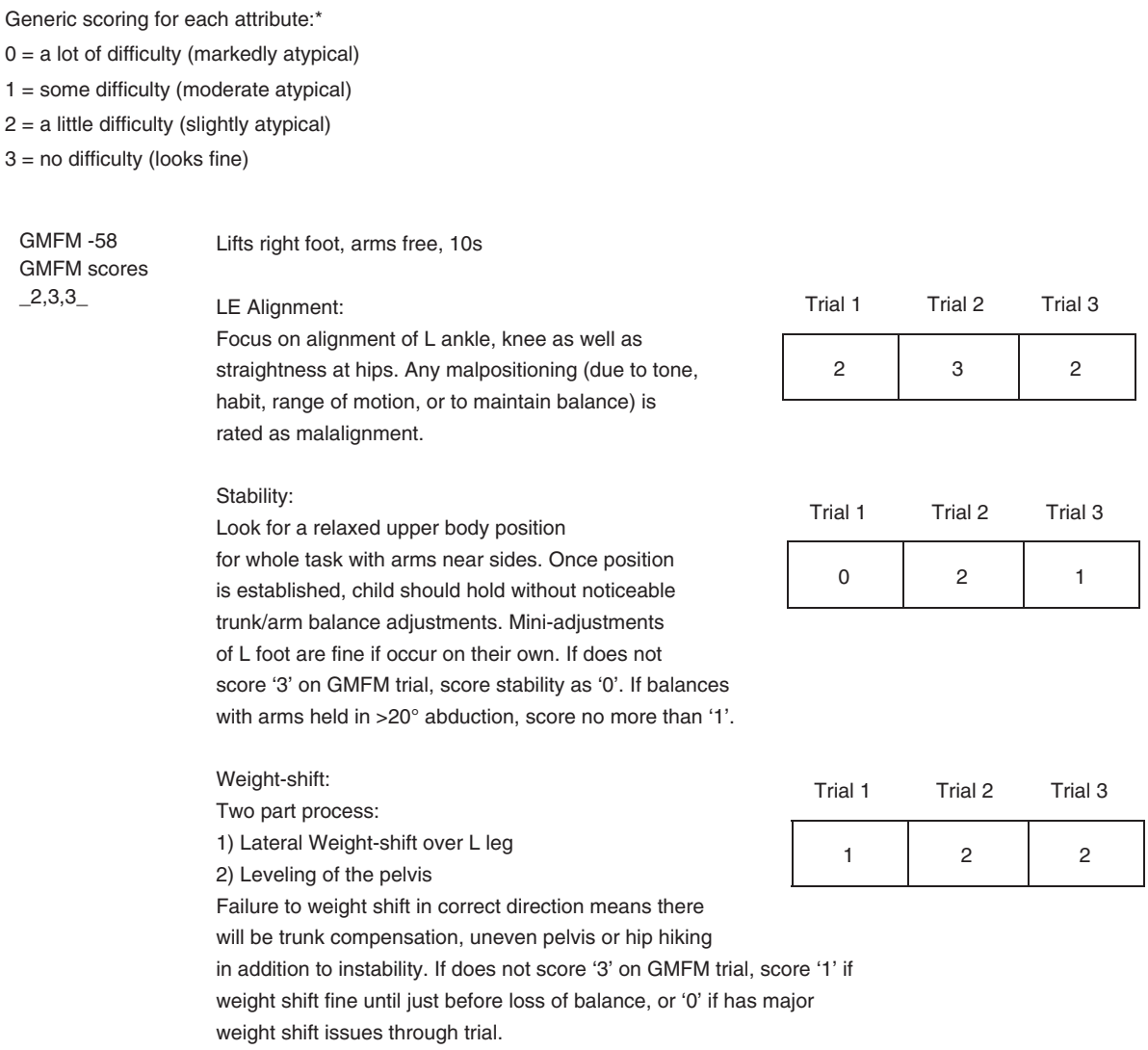

Figure 1: Sample of the Quality Function Measure (QFM) item and attribute response scales. *Details on attribute scoring descriptions and more extensive QFM materials can be obtained from www.hollandbloorview.ca and www.canchild.ca. Gross Motor Function Measure (GMFM); LE, lower extremity.

Assessors

One participating paediatric centre had two physiotherapist assessors and the other two centres had three assessors, forming a group of eight physiotherapist assessors in total. They were assigned in pairs from their centre to independently rate the QFM reliability videos of participants. In all centres, the physiotherapist assessor who conducted the GMFM-66 always scored the QFM video for the intrarater and test–retest evaluation. In the centre with two raters, the other physiotherapist assessor also scored the video for the interrater comparison. For interrater reliability in the two centres that had three physiotherapist assessors, the second physiotherapist who would rate the video was chosen at random (coin toss) by FVW. This process resulted in seven different pairs of physiotherapist assessors for the interrater reliability evaluation. Each of the physiotherapists scored 6 to 13 reliability videos in total, and the various physiotherapists pairs scored 3 to 6 reliability videos in total. For the discriminant validity evaluation sample, the physiotherapist assessor who conducted the GMFM-66 rated the child's QFM video.

Instrumentation

An example of the QFM's structure and scoring is presented in Figure 1, with more detailed scoring examples and QFM training and criterion testing information posted at www.hollandbloorview.ca and www.canchild.ca. Three trials (one 'natural' and two cued) were performed for each GMFM item that the child could undertake. The QFM was scored later from review of the video of the GMFM assessment. For each GMFM item, the score for each associated QFM attribute was calculated as the mean of these three performance trials. The natural performance trial is expected to most closely reflect the child's typical style of

performance, while the two cued performance trials allow the child to show their best performance when they are focusing on targeted aspects of quality. The resulting double weighting given to the cued performance style within these mean scores allowed extra consideration of the child's best performance (i.e. their 'capacity') in the QFM score. This is a desirable scoring feature that enables clinicians to have an idea of the child's potential to enhance the quality of their function. Finally, QFM attribute summary scores were calculated as the sum of the attribute mean scores across items, converted to percentage scores to adjust for the different number of items comprising each of the five QFM attributes.

Testing procedure

Each reliability participant had a GMFM-66 assessment (consisting of Stand and Walk/Run/Jump items 52–88) by a physiotherapist assessor at baseline and another assessment 2 weeks later, a period during which no skill changes were expected. For the discriminant validity sample, GMFM-66 baseline barefoot assessments were used. A research assistant filmed each assessment using our standard QFM video angle protocol. The GMFM video was edited by the study's primary research assistant at Holland Bloorview and finalized by FVW to remove extraneous footage. Interrater reliability involved two assessors independently scoring the baseline videos. Intrarater reliability was tested by having one assessor score the baseline video twice, and this same assessor also scored the child's retest video for the test–retest evaluation. There were at least 2 weeks and a minimum of two other children's videos assessed between scoring of a child's QFM baseline/retest videos. Assessors did not have access to previous scores.

Statistical analysis

Descriptive statistics were calculated for each attribute for the total sample and by GMFCS level. Attribute score reliabilities (intrarater, interrater, and test–retest) were evaluated using ICCs (type 2:1, a two-way random effects single measures model of absolute agreement) with associated 95% confidence intervals (CIs).[16] The ICC target was 0.80 (excellent reliability) with lower CI limits ≥0.60 (moderate reliability). Standard error of measurement (SEM) was calculated for each reliability test situation. Two other evaluations of reliability were conducted: coefficient of variation (CoV) of method error (ME)[16] with a goal of CoV less than 10%,[17] and Bland–Altman plot[18] and limits of agreement (LoA_{90}).[17]

To determine minimal detectable change (MDC) we used only the test–retest data in order to ensure that the child's session-to-session variability was considered. The decision on the confidence level for an MDC is based on how much risk a clinician is willing to take in the particular clinical circumstance with respect to misinterpreting an observed change score.[16] For this study, 80% and 90% CIs were chosen reflecting two different (albeit relatively low) degrees of risk of misinterpretation of change scores.[16,19]

Discriminant validity analyses assessed whether the QFM could detect hypothesized movement quality differences across GMFCS levels (i.e. lower quality of movement scores for all attributes in association with greater gait impairments). Strong discriminant validity is an important foundation for detection of change.[20] GMFCS levels I to III capture the span of walking ability of children with CP such that those in GFMCS level I have difficulties predominantly with running and jumping activities, those in level II are able to walk independently but with difficulty, while children in level III require an assistive device to walk. Given the very distinct walking performance differences among the three categories, the GMFCS is well suited to use as a differentiating standard in evaluation of discriminant validity. Each QFM attribute was evaluated across the three GMFCS levels using one-way analysis of variance plus Tukey's HSD test (overall alpha reduced to 0.01 to handle the five correlated attributes). Correlations between attributes were evaluated using Pearson's correlation coefficients (r), p=0.01, and $r \geq 0.80$ targeted a priori as indicative of strong association between variables. This would let us see whether, in assessing the effects of interventions, we need to measure all five attributes. If attributes are strongly correlated it might be justifiable to recommend a QFM assessing fewer attributes. However, this would only be done if the loss of a quality attribute would not be shown to compromise the content of the QFM from a clinical interpretation or goal-setting viewpoint. Since important systematic differences in attribute scores might be present even with high inter-attribute correlations, attribute scores were also compared with each other within GMFCS levels using paired t-tests and alpha adjusted to 0.001. Post hoc calculation of the study power was planned for those comparisons in which inferential analysis indicated no difference between attributes despite mean score differences of at least 3.5 points (a magnitude of potential clinical importance, i.e. similar to GMFM-66).[5] Analyses were done in MINITAB, version 15.0 (Minitab Inc., State College, PA, USA) and MEDCALC statistical software, version 12 (MedCalc, Software, Ostend Belgium), and PASS 2000 statistical software.[21]

With two observations per analysis, a sample of 30 children was sufficient to test a hypothesized ICC of 0.80 with an ICC of 0.55 as the lower acceptable limit (α=0.05, β=0.20).[22] Samples in GMPM reliability studies varied from 25 to 36 participants.[7–9] The discriminant validity sample size plan was to work with the responsiveness study sample (i.e. targeting at least 25 children in each GMFCS level). With α=0.01, this sample size would support detection of a score difference of 10.0 points (within level SD 20.0) between GMFCS levels at a power of 0.80 (β=0.20).[21]

RESULTS

Reliability analysis

Thirty-three children with CP (17 females, 16 males; mean age 8y 11mo, SD 3y 1mo) in GMFCS level I (n=17), II (n=7), and III (n=9) were enrolled for reliability analysis.

Thirty children (mean age 8y 10mo, SD 3y 3mo) returned for the retest assessment (test–retest reliability). Three children (two in GMFCS level I, one in GMFCS level III) were unable to return within the required retest interval. The interrater analysis had 32 children as the video file of one other child was damaged before review by the second rater.

Eight raters completed the replicate observations within their respective centre. All raters were introduced to the QFM and its administration at a 1-day group training session. All had at least 2 years of experience in GMFM administration. Training consisted of a review of the QFM manual/score sheet, and group scoring of QFM training videos developed by the investigative team (UB, PR, FVW). Assessors independently practised scoring with three videos (one child in each of GMFCS levels I, II, and III), and were tested on a QFM video-test to a criterion level (weighted kappa >0.80 for agreement with the first author).

For the main reliability sample (n=33), the mean GMFM-66 score was 75.6 (SD 12.4). QFM attribute mean scores varied from 45.2% (SD 27.2) for Stability to 56.1% (SD 27.5) for Alignment (Table SII, online supporting information).

Interrater, intrarater, and test–retest reliability estimates all were excellent (ICCs 0.89–0.97; Table I). The lower limit of the 95% CI for the ICCs was at least 0.83. CoV estimates were 6 to 14.6% across the reliability situations and attributes, with the lowest CoVs for the intrarater evaluation and highest for interrater. Bland–Altman plots did not reveal systematic differences in rater agreement across the range of QFM scores, and test–retest differences for the participants were within the calculated LoA for more than 90% of the time for each attribute. MDC_{80} and MDC_{90} were approximately 9 and 11 points (out of 100) respectively, across attributes, except Alignment in which it was 13.5 and 17.4 points respectively (Table I).

Table I: Reliability statistics

Attribute	ICC	95% CI	SEM	MDC_{90}	MDC_{80}	CoV (%)
Intrarater[a] (n=33)						
Alignment	0.97	0.95–0.99	4.21	–		6.7
Co-ordination	0.98	0.96–0.99	3.56	–		6.0
Dissociated movement	0.98	0.97–0.99	4.04	–		6.9
Stability	0.98	0.96–0.99	4.05	–		13.0
Weight-shift	0.97	0.95–0.97	3.78	–		7.4
Interrater[b] (n=32)						
Alignment	0.93	0.88–0.97	6.54	–		10.8
Co-ordination	0.94	0.88–0.97	6.12	–		11.7
Dissociated movement	0.92	0.85–0.96	8.12	–		14.6
Stability	0.96	0.92–0.98	5.72	–		14.2
Weight-shift	0.89	0.80–0.95	7.46	–		14.4
Test–retest[a] (n=30)						
Alignment	0.90	0.79–0.96	7.46	17.4	13.5	13.8
Co-ordination	0.96	0.92–0.96	4.80	11.2	8.7	8.0
Dissociated movement	0.97	0.94–0.99	4.65	10.8	8.4	7.6
Stability	0.96	0.94–0.99	5.45	12.7	9.9	12.7
Weight-shift	0.95	0.90–0.98	4.63	10.8	8.4	8.6

[a]All seven assessors had opportunity to do these repeat ratings with three to eight participants. [b]Two assessors (pair); all seven assessors were involved in an assessment pair with three to six participants. CI, confidence interval; CoV, coefficient of variation; ICC, intraclass correlation coefficients; MDC, minimal detectable change; SEM, standard error of measurement. The descriptive statistics for these data may be found in Table SII (online supporting information).

Validity analysis

The discriminant validity sample consisted of 112 children (49 females, 63 males; mean age 8y 2mo, SD 3y 2mo) in GMFCS levels I (n=38), II (n=27), and III (n=47). The QFM discriminated across levels (maximum $p<0.0001$) for all attribute comparisons (see Table II for mean scores). Stability mean scores were lower than Alignment, Co-ordination, and Dissociated Movement within GMFCS groups and for the total sample (maximum $p<0.01$; Table II). Stability scores were also lower than Weight-shift for all comparisons (maximum $p<0.01$), except in GMFCS level I (mean difference=0.83, p=0.56).

QFM inter-attribute correlations were very strong ($0.82<r<0.96$) for the sample overall (n=112; Table III). In the GMFCS subgroups, associations between other attributes and Alignment were the least consistent (Table III).

The mean time to score the QFM from video was 66.7 minutes (SD 33.4, minimum=15min for a child in GMFCS level III who performed few GMFM items, and maximum=180min with two children in GMFCS level II who did all GMFM items).

DISCUSSION

This study was undertaken to create a useful, clinically appropriate way to assess the quality of motor development of children with CP. It builds on the GMPM and links with the modern interval-level GMFM-66.

Reliability was excellent for all QFM attributes (ICCs≥0.89) and, with the exception of Alignment, achieved the pre-set minimum targets for measurement acceptability for ICCs, CoVs, and Bland–Altman plots. Though still excellent, the ICC for Alignment was lower for the test–retest assessment than intrarater assessment (0.90 vs 0.97), suggesting that variation in postural range of motion from one test session to the next adversely affected reliability. The mid-range mean scores for all attributes signify potential for detecting change. MDC estimates of 9 to 12 points of change were inside the 20 to 30% MDC (change in relation to total score) that has been observed and recommended as acceptable for physical rehabilitation measures.[23]

Use of several physiotherapist assessor teams may have led to more conservative estimates of reliability and MDC, but this approach enhances generalizability to the clinical environment to a population of physiotherapist raters from which this group was sampled.[24] ICC estimates generally exceeded those from GMPM reliability work with similar samples of children with CP.[7–9] This is an important outcome of the redevelopment of QFMs, as we had hoped that by following the recommendations of the GMPM developers, particularly with respect to operationalization

Table II: Discriminant validity results within GMFCS levels and for total sample (*n*=112)

	Alignment	Co-ordination	Dissociated movement	Stability	Weight-shift
GMFCS level I (*n*=38)	74.6[a1,b1] (18.5)	74.4[a2,b2] (12.0)	73.5[a3,b3] (14.3)	66.2[a1,a2,a3] (13.6)	67.08[b1,b2,b3] (12.9)
GMFCS level II (*n*=27)	50.5[c1] (19.1)	48.2[c2] (15.8)	45.6[c3] (15.7)	38.5[c1,c2,c3,c4] (15.6)	46.9[c4] (11.2)
GMFCS level III (*n*=46)	20.8[d1] (21.5)	19.5[d2] (15.0)	19.7[d3] (20.8)	9.3[d1,d2,d3,d4] (11.8)	22.4[d4] (16.2)
Total group (*n*=112)	46.7[e1] (30.7)	45.3[e2] (27.7)	44.4[e3] (29.1)	35.9[e1,e2,e3,f1] (28.1)	43.7[f1] (23.9)

Scores are presented as mean (SD). Superscripts indicate significant *p* values for pair-wise comparisons across Gross Motor Function Classification System (GMFCS) levels or for the total group, with alpha adjusted to 0.001. If the same superscript appears next to two of the mean scores in the same row of the table, this denotes that there was a significant difference between those mean scores for the group of children represented in the row. For example, superscript a1 denotes significant differences between Alignment and Stability in GMFCS level I, whereas b1 denotes significant differences between Alignment and Weight-shift in GMFCS level I. Superscript a3 denotes significant differences between Dissociated movement and Stability in post hoc power calculations were to be conducted for 'no difference' results of >3.5 points between attributes (specified a priori as a potentially clinically meaningful difference). This calculation analysis revealed that for the Alignment versus Dissociated Movement 'no difference' result in GMFCS level II (score of 50.5 vs 45.6 respectively), study power=0.44. None of the other non-significant attribute comparisons achieved differences >3.5 points, hence no further post hoc power analyses were performed.

Table III: Correlation between attributes (for total sample [*n*=112])

	Alignment	Co-ordination	Dissociated movement	Stability	Weight-shift
Alignment	–	–	–	–	–
Co-ordination	0.85[a]	–	–	–	–
Dissociated movement	0.83	0.91	–	–	–
Stability	0.84	0.95	0.88	–	–
Weight-shift	0.85	0.95	0.92	0.91	–
GMFCS level I (*n*=38)					
Alignment	–	–	–	–	–
Co-ordination	0.68	–	–	–	–
Dissociated movement	0.66	0.77	–	–	–
Stability	0.66	0.82	0.72	–	–
Weight-shift	0.74	0.91	0.81	0.79	–
GMFCS level II (*n*=27)					
Alignment	–	–	–	–	–
Co-ordination	0.29 (p=0.14)	–	–	–	–
Dissociated movement	0.32 (p=0.10)	0.80	–	–	–
Stability	0.31 (p=0.10)	0.83	0.74	–	–
Weight-shift	0.21 (p=0.28)	0.84	0.62 (p<0.01)	0.66 (p<0.01)	–
GMFCS level III (*n*=47)					
Alignment	–	–	–	–	–
Co-ordination	0.70	–	–	–	–
Dissociated movement	0.57 (p<0.05)	0.71	–	–	–
Stability	0.68 (p<0.01)	0.80	0.57 (p<0.05)	–	–
Weight-shift	0.66	0.81	0.81	0.75	–

Each cell contains the Pearson's *r*, as well as the *p* value if $p \geq 0.001$. [a]$p<0.001$ for all correlations (*r*) except those for which *p* values are displayed.

of items/rating criteria, the QFM would enhance clinicians' ability to measure movement quality.

The QFM discriminated well among children by GMFCS level. This was a key validity requirement for the QFM since, on clinical observation, movement differences related to speed, alignment, and stability, in particular, are very noticeable across GMFCS levels and may in fact underlie these functional gait distinctions. The QFM allowed us to increase the magnification of the microscope on movement quality and quantitatively proceed beyond the Gestalt style ratings that clinicians typically use when trying to make judgments about movement abilities.

The lower associations between Alignment and the other quality attributes for children in GMFCS level II might suggest that children's movement strengths (e.g. higher scores in Stability and Co-ordination) can withstand the negative impact of contractures that are in the moderate range (i.e. Alignment score of 40–50%).

The strong correlation among attributes, except Alignment, made us initially question whether all are needed. While their inclusion makes the QFM scoring longer, two key clinically relevant benefits are proposed. Firstly, use of all five attributes allowed us to uncover greater limitations related to Stability and Weight-shift in this sample, as shown by the significantly lower scores than the other three attributes. Since correlation reflects association and not agreement between scores, if there is value in knowing about specific attributes for purposes of goal setting, intervention, and outcome evaluation, all should be measured rather than inferred. The example in Figure 2 of three children, randomly selected from each GMFCS level,

illustrates how individual attribute scores might help in the goal-setting process to determine intervention focus.

Second, we do not know yet if change in one area is associated with change in another. It is evident from related work with ambulatory children with CP that this assumption may be faulty.[25] For example, if Alignment and Dissociated movement improve (perhaps after botulinum toxin injections), there still may not be gains in Stability. The current five attributes in the QFM provide an opportunity to study relative change in attribute scores, as we are doing in our current QFM responsiveness study.

One challenge to QFM score calculation for children in GMFCS level III relates to the limited number of GMFM Stand/Walk skills performed (i.e. usually 12–15 of the 37 items). Any item that the child could not initiate (i.e. GMFM score of '0') was also scored as '0' (great amount of difficulty) for each of its three QFM attributes. This approach is more conservative than removing these items from the QFM scoring equation and basing QFM scores on what was performed. It ensures that the child's relative scores align with children in GMFCS level I or II (who typically initiate all items) and keeps the scoring denominator constant for follow-up comparisons. To increase the number of skills for quality assessment in level III, three Stand dimension items (squat, stand to sit on floor, and pick up object from floor) were adapted to

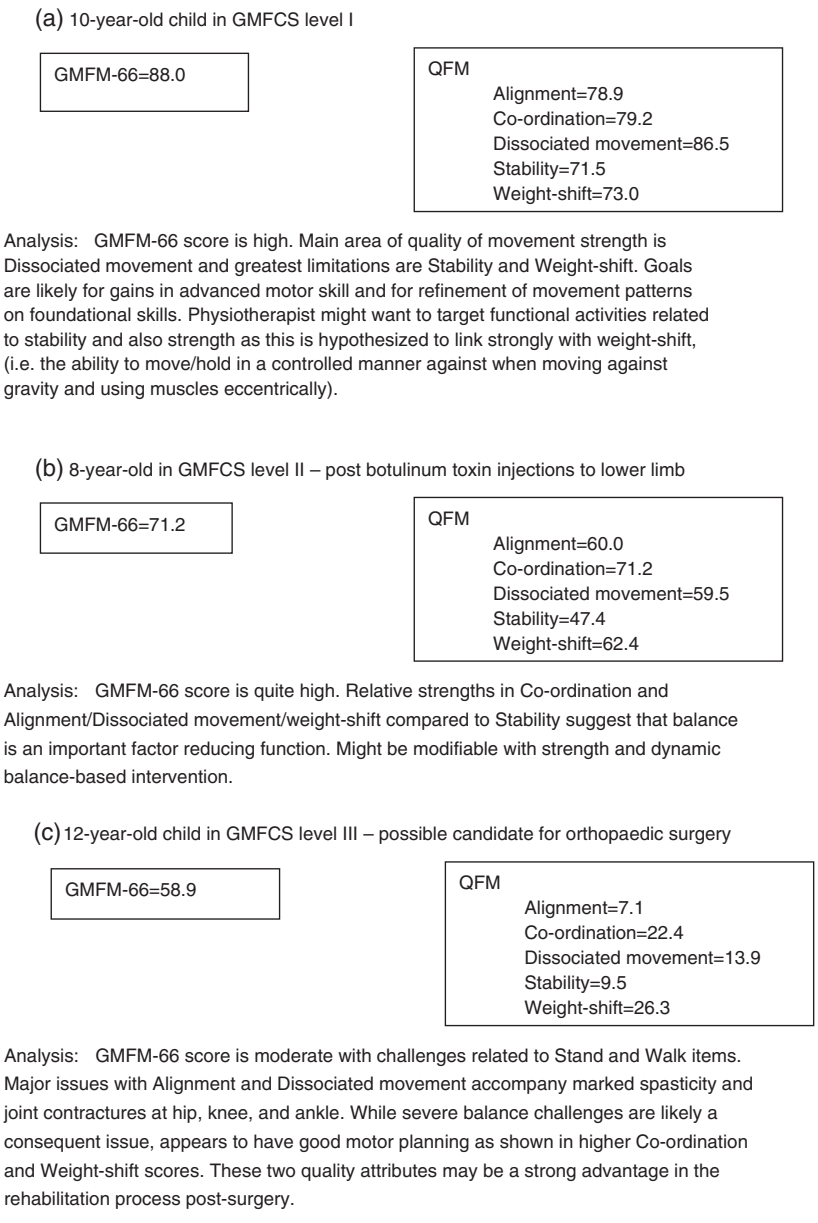

Figure 2: Quality Function Measure (QFM) profiles of a child in Gross Motor Function Classification System levels I, II, and III. GMFM, Gross Motor Function Measure.

permit the child to use hand support. While they still scored '0' on the GMFM-66 and an automatic '0' on Stability for these skills (since they held on), a full range of scores for the other two attributes linked with these items could be considered (e.g. the child might score '2' in Co-ordination reflecting reasonable speed/smoothness while holding on). This sets the stage for appropriate evaluation of movement quality if the child later progresses to hands-free ability.

The time required for QFM video-scoring was typically 60 to 90 minutes plus the administration time needed to have the child do three trials of each GMFM-66 item for the video. The benefit is that children do not perform an extra test of quality of movement. Video-rating time appeared to depend on the number of GMFM-66 items performed, complexity of the child's movement patterns, and the assessor's familiarity and comfort with the QFM. While substantial, this time parallels that for other detailed motor tests such as GMFM-66, Bruininks-Oseretsky Test of Motor Proficiency,[26] and Assisting Hand Assessment.[27]

At present, the QFM may serve best as a research tool, given the need to score from video and for physiotherapists to complete a workshop/web-based training and criterion-test process. However, this does not preclude use by clinical physiotherapists who are in environments that can support the training and video recording requirements. Since reliability was evaluated for each attribute, it may be possible to shorten the measure by selecting key attributes that align with intervention goals. This is the focus of further measurement work currently underway by our group.

One step not included in the QFM time reporting was computation of QFM scores, as this was not done by our assessors. An excel spread sheet was created that permits calculation of summary scores within 20 minutes.

Study limitations

Reliability was assessed with a sample that encompassed the three GMFCS levels to which the QFM applies. While reliability would be lower within individual GMFCS levels (given the smaller range of QFM scores), it is not possible to know the extent to which the ICC might be lowered as sample sizes within each GMFCS level were insufficient to support separate reliability analyses. This broad look at reliability is similar to reliability work with GMFM and GMPM, and forms the necessary foundation for future work on the evaluation of reliability and MDC within GMFCS levels. While this study provided a general estimate of time requirements to complete the QFM, a larger multicentre study is required to determine the associations among time and GMFCS level, QFM score, and the physiotherapist assessor's QFM experience.

Future directions

Our team is currently evaluating the responsiveness of the QFM to change in the context of botulinum toxin injections, orthopaedic surgery, conductive education intervention, and ankle-foot orthosis/barefoot comparisons. Within that work, we are exploring for the first time the relationships between acquisition of new skills (GMFM change scores) and quality of movement refinements (QFM change scores) within two measures that are directly linked. We hope that this work, taking account of both what a child does and how they do it, will contribute to a broader understanding of motor skill changes (and ultimately participation) in ambulatory children with CP. The Stability attribute of the QFM may help to address the gap that has been identified in the measurement of balance,[28] a critical area of challenge with functional consequences for children with CP. Further, we anticipate that it will support more advanced thinking about underlying readiness for change[29] and when to focus on quality of movement versus skill acquisition (acquisition of basic motor abilities)[2] in paediatric rehabilitation interventions.

ACKNOWLEDGEMENTS

We are very grateful to the children and parents who gave their time to participate in the study. We also give sincere thanks to the physiotherapists (Deana Mercier, Janet Mannen, Ginny Pearce – McMaster Children's Hospital; Keiri Porter, Rosemary Perlman, Michelle Hand – ErinoakKids Centre for Treatment and Development, Missisauga, Ontario; Linda Patrick, Blythe Dalziel, Ellen Leung, Elizabeth Hosaki at Holland Bloorview Kids Rehabilitation Hospital, Toronto), and our dedicated research assistants (Pauline Martell – ErinoakKids; Janice Penney, Nancy Goldie and the volunteers who helped with the videotaping – McMaster Children's Hospital; Susan Cohen, Gloria Lee and Celeste Meschino – Bloorview Research Institute). This research was supported by Physician Services Incorporated Foundation, Ontario, Canada. Dr Wright currently holds the Holland Bloorview Kids Rehabilitation Hospital Foundation Chair in Pediatric Rehabilitation.

The authors have stated that they had no interests that might be perceived as posting a conflict or bias.

SUPPORTING INFORMATION

The following additional material may be found online:

Table SI: Quality of movement measures in paediatrics.

Table SII: Descriptive statistics for reliability data.

REFERENCES

1. Rosenbaum P, Paneth N, Leviton A, et al. The definition and classification of cerebral palsy. *Dev Med Child Neurol* 2007; **49**: 8–14.
2. Sorsdahl AB, Moe-Nilssen R, Kaale HK, Rieber J, Strand LI. Change in basic motor abilities, quality of movement and everyday activities following intensive, goal-directed, activity-focused physiotherapy in a group setting for children with cerebral palsy. *BMC Pediatr* 2010; **10**: 26.
3. Boyce WF, Gowland C, Hardy S, et al. Development of a quality-of-movement measure for children with cerebral palsy. *Phys Ther* 1991; **71**: 820–8; discussion 28–32.

4. Pomeroy VM, Pramanik A, Sykes L, Richards J, Hill E. Agreement between physiotherapists on quality of movement rated via videotape. *Clin Rehabil* 2003; **17**: 264–72.
5. Russell DJ, Rosenbaum PL, Avery LM, Lane M. Gross Motor Function Measure (GMFM-66 and GMFM-88) User's Manual. London, UK: Mac keith press, 2002.
6. Boyce W, Gowland C, Russell D, Goldsmith C. Consensus methodology in the Development and Content Validation of a Gross Motor Performance Measure. *Physiother Can* 1993; **45**: 94.
7. Gowland C, Boyce WF, Wright V, Russell DJ, Goldsmith CH, Rosenbaum PL. Reliability of the Gross Motor Performance Measure. *Phys Ther* 1995; **75**: 597–602.
8. Thomas SS, Buckon CE, Phillips DS, Aiona MD, Sussman MD. Interobserver reliability of the gross motor performance measure: preliminary results. *Dev Med Child Neurol* 2001; **43**: 97–102.
9. Sorsdahl AB, Moe-Nilssen R, Strand LI. Observer reliability of the Gross Motor Performance Measure and the Quality of Upper Extremity Skills Test, based on video recordings. *Dev Med Child Neurol* 2008; **50**: 146–51.
10. Boyce WF, Gowland C, Rosenbaum PL, et al. The Gross Motor Performance Measure: validity and responsiveness of a measure of quality of movement. *Phys Ther* 1995; **75**: 603–13.
11. Buckon CE, Thomas SS, Jakobson-Huston S, Moor M, Sussman M, Aiona M. Comparison of three ankle-foot orthosis configurations for children with spastic diplegia. *Dev Med Child Neurol* 2004; **46**: 590–8.
12. Buckon CE, Thomas SS, Piatt JH Jr, Aiona MD, Sussman MD. Selective dorsal rhizotomy versus orthopedic surgery: a multidimensional assessment of outcome efficacy. *Arch Phys Med Rehabil* 2004; **85**: 457–65.
13. Palisano R, Rosenbaum P, Walter S, Russell D, Wood E, Galuppi B. Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Dev Med Child Neurol* 1997; **39**: 214–23.
14. Sakzewski L, Ziviani J, Van Eldik N. Test/retest reliability and inter-rater agreement of the Quality of Upper Extremities Skills Test (QUEST) for older children with acquired brain injuries. *Phys Occup Ther Pediatr* 2001; **21**: 59–67.
15. Bourke-Taylor H. Melbourne Assessment of Unilateral Upper Limb Function: construct validity and correlation with the Pediatric Evaluation of Disability Inventory. *Dev Med Child Neurol* 2003; **45**: 92–6.
16. Portney LG, Watkins MP. Foundations of Clinical Research: Applications to Practice. Upper Saddle River, NJ: Prentice Hall, 2003.
17. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med* 1998; **26**: 217–38.
18. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; **1**: 307–10.
19. Johansson GM, Hager CK. Measurement properties of the Motor Evaluation Scale for Upper Extremity in Stroke patients (MESUPES). *Disabil Rehabil* 2012; **34**: 288–94.
20. Bagley AM, Gorton G, Oeffinger D, et al. Outcome assessments in children with cerebral palsy, part II: discriminatory ability of outcome tools. *Dev Med Child Neurol* 2007; **49**: 181–6.
21. Hintze J. NCSS and PASS Statistical Software. Kaysville, Utah: Number Cruncher Statistical Systems, 2001.
22. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Stat Med* 1998; **17**: 101–10.
23. Finch E, Brooks D, Stratford PW, Mayo NE. Physical Rehabilitation Outcome Measures: A Guide to Enhanced Clinical Decision Making. Baltimore, MD: Lippincott Williams & Wilkins, 2002.
24. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Meth* 1996; **1**: 30–46.
25. Wright FV, Rosenbaum PL, Goldsmith CH, Law M, Fehlings DL. How do changes in body functions and structures, activity, and participation relate in children with cerebral palsy? *Dev Med Child Neurol* 2008; **50**: 283–9.
26. Bruininks RH, Bruininks BD. Bruininks-Oseretsky Test of Motor Proficiency. Circle Pines, MN: American Guidance Service, 2005.
27. Krumlinde-Sundholm L, Holmefur M, Kottorp A, Eliasson AC. The assisting hand assessment: current evidence of validity, reliability, and responsiveness to change. *Dev Med Child Neurol* 2007; **49**: 259–64.
28. Saether R, Helbostad JL, Riphagen II, Vik T. Clinical tools to assess balance in children and adults with cerebral palsy: a systematic review. *Dev Med Child Neurol* 2013; **55**: 988–99.
29. Darrah J, Battlett D. Dynamic systems theory and management of children with cerebral palsy: unresolved issues. *Inf Young Child* 1995; **8**: 52–9.