Minimal clinically important differences in functional motor scores in adults with spinal muscular atrophy

B. Stolte a, J.-M. Bois b, S. Bolza a, K. Kizina a, A. Totzeck a, M. Schlaga a, C. Kleinschnitz a and T. Hagenacker a

Department of Neurology, University Hospital Essen, Essen; and Institute of Medical Informatics, Biometrics and Epidemiology, University Hospital Essen, Essen, Germany

Keywords: 6MWT, HFMSE, MCID, RULM, SMA

Background and purpose: In patients with spinal muscular atrophy (SMA), functional disease scores are frequently used to evaluate the course of the disease and the efficacy of treatment. The aim of the present study was to propose minimal clinically important difference (MCID) values for motor scores in order to estimate the degree of change within a functional score that can be considered clinically meaningful.

Methods: To estimate the MCID, distribution-based approaches were used. For each assessment [Revised Upper Limb Module (RULM), Hammersmith Functional Motor Scale Expanded (HFMSE) and 6-min walk test (6MWT)] and subgroup (SMA type 2, SMA type 3, ambulatory and non-ambulatory), the following MCID values based on a cohort of 51 adults with SMA were calculated: standard error of measurement (SEm), one-half of standard deviation (1/2 SD) and one-third of standard deviation (1/3 SD) of patients’ baseline scores.

Results: For the overall cohort, the SEm, 1/2 SD and 1/3 SD MCID values were 2.9, 6.4 and 4.3 for the RULM and 4.3, 10.6 and 7.0 for the HFMSE, respectively. Subgroup analysis led to generally lower standard deviations and consecutively lower MCID values due to the significantly different motor functions of the groups. The respective MCID values for the 6MWT were 55.5 m, 71.1 m and 47.8 m.

Conclusions: Our data provide MCID values for functional motor scores commonly used in adults with SMA in order to distinguish statistical effects from ‘real’ changes. A complementary systematic consensus process could help to further adjust the MCID values we propose.

Introduction

5q-associated spinal muscular atrophy (5q-SMA) is a rare autosomal-recessive inherited neuromuscular disease that leads to progressive muscle atrophy and weakness due to degeneration of the anterior horn cells [1]. The ‘classic’ clinical classification of the disease is based on the achievement of motor milestones during childhood, with a broad range of phenotypes from severely affected infants never achieving the ability to sit [spinal muscular atrophy (SMA) type 1], to individuals able to sit but never achieving the ability to walk (SMA type 2), to ambulatory patients (SMA type 3) [2]. Because of the progressive nature of the disease, with risk of losing milestones already achieved, and as a result of the new therapeutic options leading to novel phenotypes (e.g. people with SMA type 2 capable of walking), a new classification has recently been proposed, whereby adults with SMA patients are differentiated according to their current functional status (‘non-sitter’, ‘sitter’ and ‘walker’) rather than their type [3,4]. 5q-SMA is caused by a homozygous deletion or mutation in the survival motor neuron 1 gene (SMN1) localized on...
chromosome 5q, which results in insufficient levels of SMN proteins [5,6]. Upregulating the levels of functional SMN proteins is the main target of therapeutic approaches.

Recently, huge progress has been made in SMA therapies. Nusinersen is an antisense oligonucleotide that is capable of increasing SMN protein production. In clinical trials, intrathecal treatment with nusinersen led to significantly better motor development, motor function and survival in infants and children with SMA types 1 or 2 as compared to placebo groups [7,8]. The safety and efficacy of nusinersen has also been demonstrated in adults with SMA [9]. Onasemnogene abeparvovec-xioi is a viral vector-mediated gene replacement therapy. In a cohort of 15 patients with SMA type 1, a single intravenous infusion of onasemnogene abeparvovec-xioi resulted in longer survival and improvement of motor functions when compared to historical cohorts [10].

The Hammersmith Functional Motor Scale Expanded (HFMSE) has recently been shown to be a key outcome measure in SMA studies due to the high content validity and clinical significance of the items included in the scale [11]. Further functional scores, such as the revised upper limb module (RULM) and the 6-min walk test (6MWT), are widely used to assess disease severity, clinical course and therapeutic efficacy [12–14]. However, it remains unclear what extent of change in one direction or another within a score could be considered clinically meaningful. Therefore, determining the minimal clinically important difference (MCID) values could help to evaluate treatment efficacy. The term ‘MCID’ is defined as the minimal amount of change in a score that is relevant to the patient. Three approaches have been evaluated to determine the MCID: (1) anchor-based methods, which compare changes in scores with external patient-based (i.e. patient-reported outcomes) or clinical (e.g. physiological or laboratory measures) indicators for reference [15]; (2) distribution-based methods, such as determining one-half of standard deviation (1/2 SD), one-third of standard deviation (1/3 SD) or standard error of measurement (SEm), which attempt to estimate the MCID based on the distribution of measured values in a sample [15–19]; and (3) the Delphi method, which tries to find consensus by converging opinions through repeatedly submitting a questionnaire to experts containing the respective previous results [20].

Data on the MCID in functional scores, such as the RULM, the HFMSE and the 6MWT, are lacking for patients with SMA, however, these data may be valuable in evaluating the efficacy of SMA treatments. In the present study, we used distribution-based methods to propose MCID values for adults with SMA.

Participants and methods

Participant recruitment, data collection and statistical analyses

Between July 2017 and April 2019, a total of 51 therapy-naive adults with SMA types 2 or 3 were examined in the Department of Neurology of the University Hospital Essen, Germany. The following patient data were collected: age; gender; SMA type; genetic data, including SMN2 gene copy number; age at symptom onset; age at genetic diagnosis; history of back surgery; and medication. In addition, clinical findings, such as walking ability, need for respiratory or nutritional support, and spirometry, were evaluated. Assessments included the RULM, HFMSE and 6MWT in ambulatory patients.

The study was approved by the Ethics Committee of the University Duisburg-Essen, Germany (approval number: 18-8071-BO). Written informed consent was obtained from every patient prior to study inclusion.

Statistical analyses were performed using SAS version 9.4.

Revised Upper Limb Module

The RULM is a revised version of the Upper Limb Module (ULM), specifically designed to assess upper limb function in a wide range of patients with SMA types 2 and 3. It consists of 19 scorable items, graded on a three-point system (with the exception of one item), with a maximum total score of 37 points [12]. Higher scores indicate better upper limb function. A high intra-rater reliability [intra-class correlation coefficient (ICC) = 0.948] for the RULM has been shown by analysing 126 patients with later-onset SMA [21].

Hammersmith Functional Motor Scale Expanded

The HFMSE allows the assessment of activities of daily living in patients with SMA types 2 and 3. It consists of 33 items, each scored on a scale from 0 to 2, with a maximum total score of 66 points. Higher scores indicate better motor function. The HFMSE adds 13 clinically relevant items to its predecessor, the Hammersmith Functional Motor Scale (HFMS), in order to better include ambulatory SMA patients [13]. High test-retest or intra-rater reliability has also been demonstrated for this scale (ICC = 0.959 [21] to 0.99 [13]).
Six-min walk test

The 6MWT is widely used in cardiology and pulmonology to determine aerobic capacity and endurance. It measures the distance covered by walking a flat 25-m course over a time period of 6 min and has been shown to be valid and reliable test (test-retest reliability ICC = 0.85 [22] to 0.992 [14]) in ambulatory patients with SMA.

Calculation of MCID values

Since there is no clear evidence on which distribution-based approach is the most appropriate to estimate the MCID, several approaches were used; for each assessment (RULM, HFMSE, 6MWT) three different MCID values (SEm, 1/2 SD and 1/3 SD) of the patients’ baseline scores were calculated [15–19].

The SEm values were calculated using the following equation: 

\[ SEm = SD \cdot \sqrt{1 - r_{xx}}, \]

where \( r_{xx} \) is the test-retest reliability of the respective test. For \( r_{xx} \) values from the literature described above were used: RULM: \( r_{xx} = 0.948 \); HFMSE: \( r_{xx} = 0.959 \); 6MWT: \( r_{xx} = 0.85 \).

Finally, MCID values for subgroups (SMA type 2, SMA type 3, ambulatory and non-ambulatory) were calculated.

Results

Participant characteristics

Of a total of 51 patients, 15 (10 women and 5 men) had SMA type 2 and 36 (12 women and 24 men) had SMA type 3. The mean (range) age of SMA type 2 patients was 31.4 (18–71) years. Approximately 44% of the SMA type 3 patients had a mean (range) age of 37.6 (18–71) years. Approximately 44% of the SMA type 3 patients were able to walk, whereas, by definition, none of the SMA type 2 patients were ambulatory. Seven SMA type 2 patients and six SMA type 3 patients had spondylodesis (Table 1).

| Table 1 Patient characteristics |
|-------------------------------|
|                               | SMA type 2 | SMA type 3 | Total  |
|--------------------------------|-----------|------------|--------|
| Patients, n (%)                | 15        | 36         | 51     |
| Men                            | 5 (33.3)  | 24 (66.7)  | 29 (56.9) |
| Women                          | 10 (66.7) | 12 (33.3)  | 22 (43.1) |
| Ambulatory                     | 0 (0)     | 16 (44.4)  | 16 (31.4) |
| Spondylodesis                  | 7 (46.7)  | 6 (16.7)   | 13 (25.5) |
| Mean ± SD age, years           | 31.4 ± 11.1 | 37.6 ± 13.0 | 35.8 ± 12.7 |
| Age range, years               | 18–52     | 18–71      | 18–71  |

SMA, spinal muscular atrophy.

Distribution of functional scores

From the overall cohort, the SMA type 3 subgroup and the non-ambulatory subgroup each covered the entire range of the RULM, but with clearly different median values. Significant differences with regard to the distribution of the scores were found in the comparisons between the ambulatory versus the non-ambulatory subgroups (Mann–Whitney \( U = 686.5, P < 0.0001 \)), as well as the SMA type 3 versus the SMA type 2 subgroups \( U = 182, P < 0.0001 \); Fig. 1a).

Similar to the RULM results, the overall cohort and the SMA type 3 subgroup each covered a wide range of the HFMSE, with remarkable differences in the median values between these groups. The score distributions showed significant differences between the ambulatory and the non-ambulatory subgroups (Mann–Whitney \( U = 696, P < 0.0001 \)), as well as between the SMA type 3 and the SMA type 2 subgroups \( U = 183, P < 0.0001 \); Fig. 1b).

The 6MWT could be performed in 44.4% of the patients with SMA type 3. The distance covered had a high range of 94–600 m, with a median value of 440.5 m (Fig. 1c).

Results of MCID calculation

The calculated MCID values (SEm, 1/2 SD and 1/3 SD) for the RULM ranged from approximately 3 to 6 points for the overall cohort. A comparable result was achieved in the SMA type 3 subgroup. In the SMA type 2 subgroup, the MCID values ranged between 1.2 and 2.7 points. The ambulatory subgroup had the lowest MCID values, ranging from approximately 0.5 to 1 point. In contrast, the calculated MCID values for the non-ambulatory subgroup ranged from approximately 2 to 4 points (Fig. 2a).

Similar to the results of the RULM, the calculated MCID values for the HFMSE each covered a comparable range of approximately 4–10 points for the overall cohort and the SMA type 3 subgroup. In contrast to the RULM, the lowest MCID values were calculated for the SMA type 2 subgroup (~0.5–1 point). The MCID values of the ambulatory and non-ambulatory subgroups were very close to each other (~1.5–4 points). The detailed MCID values for the HFMSE are shown in Fig. 2b.

The calculated MCID values for the 6MWT ranged from approximately 50 to 70 m, as shown in Fig. 2c.

Remarkably, because of the high test-retest reliabilities of the RULM and HFMSE, the calculated SEm values were always lower than the 1/2 SD and 1/3 SD values. In contrast, in the 6MWT, the SEm was...
Figure 1 Box-plot diagrams showing the variability in functional scores: (a) Revised Upper Limb Module (RULM), (b) Hammersmith Functional Motor Score Expanded (HFMSE) and (c) 6-min walk test (6MWT) in the whole sample and relevant subgroups. The height of the box represents the interquartile range (IQR; the difference between the 25th and 75th percentiles). The horizontal line in the box represents the group median and the diamond represents the group mean. The upper whisker extends to the largest observed value below the upper fence (located 1.5*IQR above the 75th percentile) and the lower whisker to the smallest observed value above the lower fence (1.5*IQR below the 25th percentile). The circles located outside the box and whiskers represent individual outlier values. [Colour figure can be viewed at wileyonlinelibrary.com]
Figure 2: Illustration of minimal clinically important differences (MCIDs) in (a) the Revised Upper Limb Module (RULM), (b) the Hammersmith Functional Motor Scale Expanded (HFMSE) and (c) the 6-min walk test (6MWT). Three distribution-based approaches for MCID were calculated for each subgroup: standard error of measurement (SEm), one-half of standard deviation (1/2 SD) and one-third of standard deviation (1/3 SD). [Colour figure can be viewed at wileyonlinelibrary.com]
Discussion

For the estimation of the MCID we used the distribution-based approaches most commonly used in the literature, such as the SEM, 1/2 SD and 1/3 SD [15–19]; however, it remains unclear which approach is the most appropriate for estimating the MCID. In contrast to the other two quantities, the SEM is mathematically based not only on the standard deviation of a sample, but also on the test-retest reliability ($r_{xx}$), making the SEM sample-independent [17]. This could be a key advantage of the SEM. If the degree of improvement within a score is smaller than the respective SEM, it would probably be attributable to a measurement error rather than a real change [23]. Conversely, improvements higher than the SEM can be considered as real therapeutic success.

In both the RULM and the HFMSE, the baseline scores obtained in the overall cohort showed high standard deviations (data not shown), because of the large heterogeneity of the patients’ motor functions. Since we used distribution-based methods, the high standard deviations led to high MCID values for the RULM (2.9–6.4 points) as well as the HFMSE (4.3–10.6 points). In addition, the subgroup analysis for both scales revealed highly significant differences in the score distribution between the ambulatory and non-ambulatory patients, as well as between the SMA types 3 and 2 subgroups, indicating clinically relevant differences between these subgroups in adulthood. This is in accordance with a large cross-sectional study that investigated the motor functions of 180 patients aged 1–77.5 years with SMA types 1–4 [24]. To take into account the heterogeneity of patients, and to obtain more appropriate values, we calculated the MCID values not only for the overall cohort, but also for these subgroups. This led to the following findings. First, with the exception of the MCID values for the SMA type 3 subgroup, the calculated MCID values for the other subgroups were clearly lower when compared to the overall cohort. This applies to both the RULM and the HFMSE and indicates that the SMA type 3 subgroup was similarly heterogeneous in terms of motor functions as the overall cohort. Second, the calculated MCID values for the ambulatory subgroup in the RULM were quite low (0.4 to 0.8 points), which is due to a ceiling effect (13 of 16 ambulatory patients reached the maximum score of 37 points). This is not surprising since assessments of the upper limb were developed to capture clinically meaningful function in non-ambulant patients with low residual motor function [12,25]. Remarkably, the RULM was designed to eliminate the ceiling effect of its predecessor, the ULM, and to make it more useful in a wider SMA population [12]. Nevertheless, our observation of a ceiling effect is consistent with a longitudinal study conducted in 114 patients with SMA types 2 and 3, in which the mean score of the ambulatory subgroup was 34.2 points and approximately one-third of the ambulatory patients reached the maximum score [26]. Thus, the RULM could be used as a limited instrument to identify relevant improvements in the ambulatory subgroup. Third, the calculated MCID values for the SMA type 2 subgroup in the HFMSE were also reasonably low (0.5–1.2 points), which is attributable to a floor effect (13 of 15 SMA type 2 patients achieved a score between 0 and 4 of 66 points, with a median value of 2 points). A floor effect of the HFMSE for the SMA type 2 patients has also been observed by Wadman et al. [24]. This makes the HFMSE a limited tool for assessing weaker SMA patients [27]. Fourth, despite the highly significant differences in score distribution, the calculated MCID values for the ambulatory and non-ambulatory subgroups were quite close to each other (1.8–4.3 points and 1.5–3.8 points, respectively). However, this applies only to the HFMSE.

Data on the MCID in functional scores are lacking for patients with SMA.

For the Modified Hammersmith Functional Motor Scale (MHFMS), a variant of the HFMS, a change of 3 points was proposed to be clinically meaningful in patients with SMA type 2 [28,29]. However, this analysis did not use the HFMS, which also includes patients with SMA type 3. A study conducted in 180 patients with SMA types 1–4 demonstrated a decline of 0.5 points per year in the HFMSE score [24]. Furthermore, analysis of the natural disease course in 268 patients with SMA types 2 and 3 aged between 2.5 and 55.5 years (HFMSE mean 23.91, range 0–66) showed that more than 75% of patients had changes ± 2 points, while fewer than 10% showed an improvement of more than 2 points after 12 months [30]. Thus, changes above 2 points in the HFMSE are generally considered clinically relevant. These data are consistent with our estimation that the MCID values for the ambulatory and non-ambulatory subgroups range between 1.5 and 4.3 points. However, in patients with SMA type 2 the calculated MCID values were notably lower because of the floor effect mentioned above.

For the RULM, analysis of the natural disease course in 114 patients with SMA types 2 and 3 aged between approximately 2 and 50 years showed that the 12-month changes ranged between −7 and +9
the test-retest reliability of \( r_{xx} \) as high (data not shown). For the 6MWT, we used the HFMSE would have been only approximately half Hagen between studies [13,21]. It is noteworthy that using O

bers of participants, evaluators and study centres reliabilities are likely to be a result of different num-
SMA types 2 and 3 in 2007. The different test-retest developed and evaluated this scale on 38 patients with
Elsheikh et al
−
2 points, with the majority of patients (67%) between –2 and + 2 points. Similar to the HFMSE, therefore, in the RULM, changes beyond 2 points are generally considered clinically meaningful [26]. These data are consistent with our estimation that the MCID values for the SMA type 2 subgroup and the non-ambulatory subgroup range from 1.2 to 2.7 points and from 2 to 4 points, respectively. However, our MCID values for ambulatory patients were considerably lower because of the ceiling effect described above.

To our knowledge, sufficient data on the MCID values of the 6MWT are not available for patients with SMA.

Since functional disease scores were performed only once (further assessments were only performed in case of therapy with nusinersen), we were unable to calculate the test-retest reliabilities. Hence, a limitation of the present study could be the use of test-retest reliabilities that were obtained from the literature to estimate the respective SEM values. Importantly, the calculations of the 1/2 SD and 1/3 SD were not affected by this, as no reliability was required. Data from the literature on test-retest reliabilities for the RULM and the HFMSE are rare. For both, we used data recently published by Glanzman et al. [21]. These were collected in a global pivotal clinical trial of nusinersen conducted by the CHERISH study of 126 children aged 2–12 years with late-onset SMA (i.e. non-ambulatory SMA type 2 patients) [8]. The calculations of the test-retest reliabilities were based on the respective screening and baseline scores prior to the initiation of therapy. The assessments were performed by 61 specially trained evaluators at 26 study sites [8,21]. With regard to our calculations, the young age of the patients, and the fact that only patients with SMA type 2 were studied, could limit the usability of the reliabilities for our data. However, for the RULM, no other test-retest reliability was available in the literature, because it is a relatively new scale. For the HFMSE, a higher test-retest reliability of \( r_{xx} = 0.99 \) was postulated by O’Hagen et al. [13], who developed and evaluated this scale on 38 patients with SMA types 2 and 3 in 2007. The different test-retest reliabilities are likely to be a result of different numbers of participants, evaluators and study centres between studies [13,21]. It is noteworthy that using O’Hagen’s value for \( r_{xx} \), our calculated SEM values for the HFMSE would have been only approximately half as high (data not shown). For the 6MWT, we used the test-retest reliability of \( r_{xx} = 0.85 \) as published by Elsheikh et al. in 2019 due to the relatively large number of assessed participants (\( n = 30 \)) and comparable demographics, with a mean age of approximately 37 ± 9 years [22]. Notably, the use of another published quite high test-retest reliability (\( r_{xx} = 0.992 \)) would have resulted in a very low SEM (not shown). However, the corresponding study was performed in only 17 patients aged between 10 and 49 years [14].

The relatively small number of patients is a further limitation of the present study. On the one hand, the mean values of the baseline scores are only representative to a limited extent. On the other hand, the phenotypic heterogeneity in the overall cohort had a strong impact on the score distribution, leading to high standard deviations. At least the latter should have been mitigated by the subdivision into appropriate subgroups.

It is important to note that when evaluating the success of therapy, not only changes in motor scales should be taken into account, but also the expectations and individual conditions of the patients. Recently, it has been shown that changes should be considered relative to residual motor function and that even small changes or the prevention of disease progression can be valuable. In addition, other aspects of the disease relevant to everyday life, such as swallowing or respiratory difficulties and fatigue, are not captured by the scales, but may have an influence on quality of life [31]. Remarkably, the US Food and Drug Administration’s guidance on clinical trials therefore also requires the inclusion of patient experience data and input from patients and caregivers to define a meaningful change [32].

Finally, distribution-based approaches do not provide direct information about the MCID, but they can help to distinguish statistical effects from ‘true’ changes, especially where anchor-based estimates are not available. In general, the definitive calculation of MCID values should be based on multiple approaches and a triangulation of methods [15]. A systematic consensus process, such as the Delphi method, could therefore help to converge and further adjust the MCID values we propose for the respective assessments.

In conclusion, functional motor scores, including the RULM, HFMSE and 6MWT, can help to evaluate the course of disease and efficacy of treatment in adults with SMA. Our data provide MCID values for patients with SMA in order to better assess what degree of change within a functional score could be considered clinically meaningful or a real therapeutic success. Due to the large phenotypic heterogeneity among adults with SMA, the analysis of subgroups such as ambulatory versus non-ambulatory appears to be useful for an appropriate calculation of MCID values. Because the SEM takes into account the high reliability of the functional scores, the use of the SEM could be preferred over the other quantities. However,
the definitive calculation of MCID values should ideally be based on multiple approaches. Complementary approaches could help to further adjust the MCID values we propose.

Acknowledgements

Funding information: This research received no specific grants from any funding agency in the public, commercial or not-for-profit sectors.

Disclosure of conflict of interest

B.S. and C.K. received travel reimbursement and speaker honoraria from Biogen. K.K. received travel reimbursement from Biogen. T.H. received travel reimbursement from Biogen and speaker honoraria from Novartis and Biogen and research support from Biogen. J.M.B., S.B., A.T. and M.S. declared no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

1. Mercuri E, Bertini E, Iannaccone ST. Childhood spinal muscular atrophy: controversies and challenges. Lancet Neurol 2012; 11: 443–452.
2. Munsat TL, Davies KE. International SMA consortium meeting (26–28 June 1992, Bonn, Germany). Neuromuscul Disord 1992; 2: 423–428.
3. Sansone VA, Walter MC, Attarian S, et al. Measuring outcomes in adults with spinal muscular atrophy - challenges and future directions - meeting report. J Neuromuscul Dis 2020; Pre-print: 1–12.
4. Mercuri E, Finkel RS, Muntoni F, et al. Diagnosis and management of spinal muscular atrophy: Part I: recommendations for diagnosis, rehabilitation, orthopedic and nutritional care. Neuromuscul Disord 2018; 28: 103–115.
5. Lefebvre S, Burglen L, Reboullet S, et al. Identification and characterization of a spinal muscular atrophy-determining gene. Cell 1995; 80: 155–165.
6. Lefebvre S, Burlet P, Liu Q, et al. Correlation between severity and SMN protein level in spinal muscular atrophy. Nat Genet 1997; 16: 265–269.
7. Finkel RS, Mercuri E, Darras BT, et al. Nusinersen versus sham control in infantile-onset spinal muscular atrophy. N Engl J Med 2017; 377: 1723–1732.
8. Mercuri E, Darras BT, Chiriboga CA, et al. Nusinersen versus sham control in later-onset spinal muscular atrophy. N Engl J Med 2018; 378: 625–635.
9. Hagenacker T, Wurster CD, Günther R, et al. Nusinersen in adults with 5q spinal muscular atrophy: a non-interventional, multicentre, observational cohort study. Lancet Neurol 2020; 19: 317–325.
10. Mendell JR, Al-Zaidy S, Shell R, et al. Single-dose gene-replacement therapy for spinal muscular atrophy. N Engl J Med 2017; 377: 1713–1722.
11. Pera MC, Corrati G, Foricina N, et al. Content validity and clinical meaningfulness of the HFMSE in spinal muscular atrophy. BMC Neurol 2017; 17: 39.
12. Mazzone ES, Mayhew A, Montes J, et al. Revised upper limb module for spinal muscular atrophy: development of a new module. Muscle Nerve 2017; 55: 869–874.
13. O’Hagen JM, Glanzman AM, McDermott MP, et al. An expanded version of the hammersmith functional motor scale for SMA II and III patients. Neuromuscul Disord 2007; 17: 693–697.
14. Dunaway Young S, Montes J, Kramer SS, et al. Six-minute walk test is reliable and valid in spinal muscular atrophy. Muscle Nerve 2016; 54: 836–842.
15. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol 2008; 61: 102–109.
16. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Med Care 2003; 41: 582–592.
17. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. J Clin Epidemiol 1999; 52: 861–873.
18. Yeo F, Ng CC, Loh KWJ, et al. Minimal clinically important difference between the EORTC QLQ-CIPN20 for worsening peripheral neuropathy in patients receiving neurotoxic chemotherapy. Support Care Cancer 2019; 27: 4753–4762.
19. Ousmen A, Touraine C, Deliu N, et al. Distribution and anchor-based methods to determine the minimally important difference on patient-reported outcome questionnaires in oncology: a structured review. Health Qual Life Outcomes 2018; 16: 228.
20. McKenna HP. The Delphi technique: a worthwhile research approach for nursing? J Adv Nurs 1994; 19: 1221–1225.
21. Glanzman AM, Mazzone ES, Young SD, et al. Evaluating training and reliability for SMA global nusinersen trials1. J Neuromuscul Dis 2018; 5: 159–166.
22. Elsheikh B, King W, Peng J, et al. Outcome measures in a cohort of ambulatory adults with spinal muscular atrophy. Muscle Nerve 2019; 61: 187–191.
23. Copay AG, Subach BR, Glassman SD, Polly DW, Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. Spine J 2007; 7: 541–546.
24. Wadman RI, Wijngaarde CA, Stam M, et al. Muscle strength and motor function throughout life in a cross-sectional cohort of 180 patients with spinal muscular atrophy types 1c–4. Eur J Neurol 2018; 25: 512–518.
25. Mazzone E, Bianco F, Martinelli D, et al. Assessing upper limb function in nonambulant SMA patients: development of a new module. Neuromuscul Disord 2011; 21: 406–412.
26. Pera MC, Corrati G, Mazzone ES, et al. Revised upper limb module for spinal muscular atrophy: 12 month changes. Muscle Nerve 2019; 59: 426–430.
27. Mazzone E, De Sanctis R, Fanelli L, et al. Hammer-smith functional motor scale and motor function measure-20 in non ambulant SMA patients. *Neuromuscul Disord* 2014; 24: 347–352.

28. Swoboda KJ, Scott CB, Reyna SP, et al. Phase II open label study of valproic acid in spinal muscular atrophy. *PLoS One* 2009; 4: e5268.

29. Swoboda KJ, Scott CB, Crawford TO, et al. SMA CARNI-VAL trial part I: double-blind, randomized, placebo-controlled trial of L-carnitine and valproic acid in spinal muscular atrophy. *PLoS One* 2010; 5: e12140.

30. Mercuri E, Finkel R, Montes J, et al. Patterns of disease progression in type 2 and 3 SMA: implications for clinical trials. *Neuromuscul Disord* 2016; 26: 126–131.

31. McGraw S, Qian Y, Henne J, Jarecki J, Hobby K, Yeh WS. A qualitative study of perceptions of meaningful change in spinal muscular atrophy. *BMC Neurol* 2017; 17: 68.

32. FDA. Patient-focused drug development: collecting comprehensive and representative input. https://www.fda.gov/media/139088/download. Accessed April 22, 2020. *Guidance document.*