

Research Report

Evaluator Training and Reliability for SMA Global Nusinersen Trials¹

Allan M. Glanzman^{a,*}, Elena S. Mazzone^b, Sally Dunaway Young^{c,d}, Richard Gee^e, Kristy Rose^f, Anna Mayhew^g, Leslie Nelson^h, Chris Yunⁱ, Katie Alexander^j, Basil T. Darras^k, Zarazuela Zolkipli-Cunningham^l, Gihan Tennekoon^l, John W. Day^m, Richard S. Finkelⁿ, Eugenio Mercuri^b, Darryl C. De Vivo^c, Ron Baldwin^j, Kathie M. Bishop^j and Jacqueline Montes^{c,d}

^a*Department of Physical Therapy, Children's Hospital of Philadelphia, Philadelphia, PA, USA*

^b*Department of Neurology, Catholic University, Rome, Italy*

^c*Department of Neurology, Columbia University, New York, NY, USA*

^d*Department of Rehabilitation and Regenerative Medicine, Columbia University, New York, NY, USA*

^e*Department of Physical Therapy, Lucile Packard, Children's Hospital Stanford, CA, USA*

^f*Sydney University, Sydney, Australia*

^g*Newcastle University, Newcastle, UK*

^h*UT Southwestern Medical Center, Dallas, TX, USA*

ⁱ*Lexa Enterprises, Inc., USA*

^j*Ionis Pharmaceuticals, Carlsbad, CA, USA*

^k*Department of Neurology, Boston Children's Hospital and Harvard Medical School, Boston, MA, USA*

^l*Division of Neurology, The Children's Hospital of Philadelphia and Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA*

^m*Department of Neurology, Stanford University School of Medicine, Stanford, CA, USA*

ⁿ*Department of Neurology, Nemours Children's Hospital, Orlando, FL, USA*

Abstract.

Background: Training methodology was established to optimize reliability of outcome measures in the nusinersen clinical trials. The Children's Hospital of Philadelphia Infant Test of Neuromuscular Disorders (CHOP INTEND), Hammersmith Functional Motor Scale Expanded (HFMSE), and Revised Upper Limb (RULM) were primary or secondary outcomes.

Methods: Video review, quarterly conference calls, and item scoring checks supported evaluator competence. Baseline and screening along with video review established intra and inter-rater reliability.

Results: Inter and intra-rater reliability were both excellent. Intraclass correlation coefficients (ICC) ranged between 0.906–0.994 across initial training meetings and 0.824–0.996 across annual retraining meetings. This was similar for CHOP INTEND (ICC = 0.824–0.951), HFMSE (ICC = 0.981–0.996), and RULM (ICC = 0.966–0.990). Intra-rater reliability for the CHOP INTEND, HFMSE, and RULM were ICC = 0.895 (95%CI: 0.852–0.926; $n = 116$), ICC = 0.959 (95%CI: 0.942–0.971; $n = 125$), and ICC = 0.948 (95%CI: 0.927–0.963; $n = 126$) respectively.

¹We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

*Correspondence to: Allan M. Glanzman, The Children's Hospital of Philadelphia, 34th and Civic Center Blvd, Philadelphia,

10104, PA, USA. Tel.: +1 215 590 7730; E-mail: glanzmana@email.chop.edu.

Conclusions: Rigorous evaluator training ensures reliability of assessment of subjects with spinal muscular atrophy (SMA) in multicenter international trials.

Keywords: SMA, spinal muscular atrophy, nusinersen, outcome measures, outcome assessment, reliability of results, multicenter studies

INTRODUCTION

Spinal muscular atrophy (SMA), is a genetically determined motor neuron disease with presentation ranging from severe respiratory compromise and functional impairment to later onset proximal weakness that can impact ambulation and higher level functional motor skills [1, 2]. This phenotypic variability results from deletion of the *SMN1* gene that forces a reliance on *SMN2*, a partially functional paralog of *SMN1*. Multiple disease modifying therapies, some targeting the *SMN2* gene, are being evaluated in early and later phase trials. For later stage clinical trials, [3] functional outcome measures with established natural history data are critical both for study planning and interrogation of treatment effect. The outcomes used in the nusinersen clinical trials are well-established in SMA [1, 4, 5].

Established reliable outcomes are necessary for the success of any clinical trial. Reliability depends on consistency of administration and performance of the subject, and represents a major component of the variability that is experienced within any trial [6]. While reliability of these clinical assessments are well-established in observational trials performed at select specialized centers, study specific training and harmonization among evaluators is essential particularly in global multinational clinical trials. The importance of training for the evaluators administering outcomes in any trial cannot be over stated and provides the basis on which the reliability of functional outcome measurement data rests.

Children with SMA also pose special challenges to reliability in the context of clinical trials. The phenotype of patients with SMA demonstrates significant variability, from infants with type I SMA which presents at or shortly after birth to later onset type II and III SMA with functional skills that, respectively, span the range from never sitting to sitting or walking independently [3]. Infants and children with type I and type II SMA present additional challenges with respect to measurement that relate to fatigue, pulmonary limitations, and developmental progression [4, 7].

Nusinersen, an antisense oligonucleotide (ASO) is the first FDA approved Antisense oligonucleotide

based therapy for SMA [8]. Nusinersen increases exon 7 inclusion in *SMN2* mRNA transcripts, production of full length *SMN* protein [7, 9] and has received FDA approval for pediatric and adult patients with SMA [10]. Two, global phase III clinical trials of nusinersen began in 2014. The first included those with infantile onset SMA with symptom onset ≤ 6 months of age and the second included children with later onset SMA with symptom onset after 6 months who were able to sit independently but were never able to walk independently. Both included functional outcome measures and the attainment of motor milestones as primary or secondary endpoints [11, 12]. As such, a rigorous training methodology for clinical evaluators was needed to optimize the reliability of outcome measurement within the trials. The purpose of this report is to 1) describe the training methods and procedures necessary to standardize clinical evaluations and outcome measures in a global, multicenter, phase 3 clinical trial, and 2) to demonstrate the test-retest and inter-rater reliability of The Children's Hospital of Philadelphia Infant Test of Neuromuscular Disorders (CHOP INTEND), the Expanded Hammersmith Functional Motor Scale (HFMSE), and the Revised Upper Limb (RULM) used in these 2 pivotal SMA clinical trials.

METHODS

Training methodology

Training for the CHOP INTEND used in the infantile onset SMA trial was conducted across all 32 sites and 73 evaluators. Sites identified primary and back up evaluators. There were 15 sites in North America (32 evaluators), 10 in the Europe (23 evaluators), 2 in Japan (6 evaluators) and 5 sites in the Asia-Pacific region including Australia, Korea, Hong Kong, and Taiwan (12 evaluators). These initial training meetings in North America and Europe were followed by yearly retraining meetings. The initial training for the infantile onset trial was conducted in North America ($n = 25$) and Europe ($n = 23$) with both Asia-Pacific and the European evaluators grouped in the European training. Translation was provided for the non-English speaking European evaluators by

experienced evaluators who were bilingual or professional translators. Testing materials were translated into local languages base on evaluator need both in Europe and Asia.

Training for the HFMSE and the RULM, used in the later onset SMA trial, was conducted across all 26 sites and 61 evaluators. Many of the sites and evaluators between the infantile and later onset studies overlapped though not all. Of the 26 total sites 22 sites participated in both studies. Similar to the training for the CHOP INTEND, sites identified primary and back up evaluators. There were 13 sites (28 evaluators) in North America, 9 (23 evaluators) in Europe, 2 (6 evaluators) in Japan and 2 (4 evaluators) in the Asia-Pacific region. Investigator meetings were held in North America, Europe, and Asia and yearly retraining meetings were held for this trial as well.

Initial training was standardized and completed in conjunction with the investigator meetings over two days and included clinical trial and outcome measure background, specific outcome measure item review followed by video review. Hands on patient training, was not possible due to the stress of repetitive examination. Hands-on simulations were instead used with mannequins to go over hand placement for item administration, for the type I assessment. Finally, reliability was checked by video review. For those evaluators who could not attend the investigator meeting, or entered as evaluators after the investigator meeting, the same training modules and methodology was used for these one on one training sessions. Novice therapists had a face to face training, experienced therapists were trained via live web based conferencing. Attendance was also required at yearly retraining meetings which included inter-rater reliability assessments and followed by quarterly conference calls which included reliability checks or

item scoring practice. At each retraining, item scoring was reviewed and the prior quarter's queries were reviewed to assure consistency across trials. Individual queries between meetings were catalogued and discussed at the next meeting and augmented by video review examples. These additional support methods assured that the methodology stayed fresh in the evaluator's minds and that consistent administration occurred across centers.

Each evaluator was also asked to video tape the first screening for subjects in the later onset SMA trial. The video of this assessment was reviewed by their one on one trainer and feedback provided as needed prior to the baseline assessment. For the evaluators in the infantile onset trial, support was available to answer questions in real time or following evaluation throughout the study related to test administration. However, specific administration was not coached by video between the screening and baseline data collection points so as not to influence test administration after screening data began to be collected since this was part of the combined baseline. Evaluator training was designed to provide prospective support for the evaluators and also to be reactive to questions that arose during the trial from individual evaluators and to assure that there was consistency of methodology across the sites.

Characteristics of the clinical evaluators and trainers

Primary and back up evaluators that participated in both trials needed to have experience across the range of subjects they might encounter in the trial. Experience with the outcome measures, SMA, and pediatrics were surveyed and a specific training program (Fig. 1) was recommended. Evaluators with limited experience were prospectively identified

Initial Training Two day in person training)	Trial background & Scale development Item video review: CHOP INTEND (type I study) HFMSE, RULM and WHO MM (Type II study) Hands-on simulation Training Video Reliability
Refresher training Annual (evaluator meeting) Quarterly (webinar)	Study update Item video review Centrally curated Q&A from prior period Inter-rater video reliability (annual) Independent video scoring checks
Video review Initiation of evaluator administration	First HFMSE and 2 RULMs, feedback prior to first baseline Video reviewed for quality assurance only, not scoring

Fig. 1. Training Methodology.

and appropriately supported during the training period.

The training group consisted of 8 physical therapists all with experience in SMA and clinical trials local to each region. The group had a median of 21 years' experience (range 10–30 years) as physical therapists and 15 years (range 10–27 years) with neuromuscular disorders. The geographic representation in Asia-Pacific Region (one physical therapist), Europe (two physical therapists), and North America (five physical therapists) allowed for additional onsite training should that be needed to optimize reliability.

The evaluators all responded to a questionnaire that scored their level of experience as therapists in pediatrics, neuromuscular disease, and clinical trials. The group was sorted into advanced, intermediate, and beginner evaluators. Evaluators in the infantile onset trial had on median of 15 years of experience as therapists (25/75 IQ range: 9–25 years) and 12.5 and 9.0 years working with children and neuromuscular disease, respectively across the study. Those in the later onset trial had on median of 14 years of experience as therapists (25/75 IQ range 6–21 years) and 9.2 and 7.7 years working with children and neuromuscular disease, respectively. These factors were then used to determine the intensity and method of training that would be required should an evaluator not be able to attend one of the centralized meetings that were provided for face to face training or to provide extra support in the context of the training meeting.

Subjects

Data was collected as part of multicenter global clinical trials [13], IRB approval was secured at each site and subjects' parents/legal guardian's provided informed consent. Screening and baseline evaluations were available from 116 subjects with infantile onset SMA, 3 were excluded due to screening failures. One-hundred twenty-five subjects with later onset SMA were available for analysis on the HFMSE, 1 was excluded due to screening failure. One-hundred twenty-six subjects with later onset SMA were available for analysis on the RULM, 1 was excluded due to screening failure.

Statistical methods

Intra-class correlation coefficients (ICC) and 95% confidence intervals (CI) using a one-way random effects, single measures, analysis of variance model

was used to determine inter-rater reliability from annual training and retraining meetings and intra-rater reliability from screening assessments. Pearson correlation coefficients were used to assess the performance of individual evaluators to the gold standard determined by the consensus of the training team for all three scales.

RESULTS

Intra-rater reliability for the CHOP INTEND was evaluated using baseline and screening visits of each subject, performed at least 1 day and no more than 21 days apart. Intra-rater reliability for the CHOP INTEND was ICC (1,1)=0.895 (95%CI: 0.852–0.926; $n=116$). The inter-rater reliability by video review was ICC (1,1)=0.951 and 0.906 for each of the two initial training sessions. Comparison to the gold standard score ranged from $r=0.714$ –0.996 (Table 1).

Intra-rater reliability using HFMSE screening and baseline visits of each subject for the HFMSE was ICC (1,1)=0.959 (95%CI: 0.942–0.971; $n=125$). Initial training session inter-rater reliability by video review by region for the HFMSE in the later onset SMA patients ranged between ICC (1,1)=0.987 and 0.994. The comparison of each evaluator to what was viewed as the gold standard score on the HFMSE was nearly perfect and ranged from $r=0.987$ –1.000 (Table 2).

Intra-rater reliability for the RULM, evaluated using screening and baseline visits of each subject, was ICC (1,1)=0.948 (95%CI: 0.927–0.963; $n=126$). Initial training session inter-rater reliability for the RULM based on video assessment between evaluators ranged between ICC (1,1)=0.966–0.990. When compared to the gold standard score correlations ranged from $r=0.895$ –1.00 (Table 3).

Reliability at the follow up retraining meetings reflected the consistency found at the initial meetings (see Tables 1–3 for intra-class correlation coefficients, confidence intervals, group size, and Pearson Correlation coefficients).

DISCUSSION

Functional outcome measures exhibit excellent reliability in global phase 3 clinical trials in infants and children with SMA. Essential to the success in multi-centered, multi-national trials is a rigorous

Table 1
Children’s Hospital of Philadelphia Infant Test of Neuromuscular Disorders (CHOP INTEND)

Children’s Hospital of Philadelphia Infant Test of Neuromuscular Disorders			
Intra-Rater Reliability: ICC = 0.895 (95%CI: 0.852–0.926; n = 116)			
Inter-Rater Reliability	Trainers	Evaluators	Evaluators vs. gold standard
Initial reliability North America	ICC = 0.963, 95%CI:0.946–0.977 n = 7	ICC = 0.951 95% CI: 0.931–0.969 n = 25	r = 0.941 – 0.996 n = 25
Initial reliability Europe/Asia Pacific		ICC = 0.906 95% CI: 0.867 – 0.940 n = 23	r = 0.775 – 0.987 n = 23
Retraining North American	ICC = 0.867 95%CI:0.808–0.915 n = 5	ICC = 0.871 95% CI: 0.822–0.914 n = 23	r = 0.714 – 0.954 n = 23
Retraining Australia/Asia Pacific		ICC = 0.824 95% CI: 0.760–0.882 n = 14	r = 0.756 – 0.922 n = 14
Retraining Europe		ICC = 0.889 95% CI: 0.845–0.927 n = 20	r = 0.716 – 0.903 n = 20
Retraining Japan/Asia Pacific		ICC = 0.869 95% CI: 0.815–0.915 n = 9	r = 0.789 – 0.911 n = 9

ICC = Interclass correlation coefficient; r = Pearson correlation coefficient.

Table 2
Expanded Hammersmith Functional Motor Scale (HFMS)

Expanded Hammersmith Functional Motor Scale			
Intra-Rater Reliability: ICC = 0.959 (95%CI: 0.942–0.971;n = 125)			
Inter-Rater Reliability	Trainers	Evaluators	Evaluators vs. gold standard
Initial reliability North America	ICC = 0.997; 95%CI:0.988–1.000 N = 7	ICC = 0.994 95% CI: 0.979–1.000 n = 23	r = 0.994 – 1.000 n = 23
Initial reliability Asia Pacific		ICC = 0.987 95% CI: 0.948–1.000 n = 12	r = 0.996 – 1.000 n = 12
Initial reliability Europe		ICC = 0.991 95% CI: 0.966–1.000 n = 17	r = 0.987 – 1.000 n = 17
Retraining North America	ICC = 0.995 95%CI:0.976–1.000 n = 5	ICC = 0.993 95% CI: 0.974–1.000 n = 22	r = 0.986 – 1.000 n = 22
Retraining Europe		ICC = 0.991 95% CI:0.965–1.000 n = 12	r = 0.997 – 1.000 n = 12
Retraining Australia		ICC = 0.996 95% CI:0.981–1.000 n = 7	r = 0.999 – 1.000 n = 7
Retraining Japan		ICC = 0.981 95% CI:0.921–1.000 n = 9	r = 0.997 – 1.000 n = 9

ICC = Intraclass correlation coefficient; r = Pearson correlation coefficient.

and comprehensive training plan as well as standardized assessments of reliability of clinical evaluators. We have described the evaluator training program and established it was effective in allowing a replication of the reliability experienced during the test development process [5, 14, 15]. This result was sup-

ported by face-to-face training experiences, quarterly conference calls, and annual retraining of the evaluator team. The collection of reliability data allowed the team to confirm the consistency of administration both across and within evaluators through the duration of the study.

Table 3
Revised Upper Limb Module (RULM)

Revised Upper Limb Module			
Intra-Rater Reliability: ICC = 0.948 (95%CI: 0.927–0.963;n = 126)			
Inter-Rater Reliability	Trainers	Evaluators	Evaluators vs. gold standard
Initial reliability North America	ICC = 0.994 95%CI: 0.975–1.000 n = 7	ICC = 0.990 95% CI: 0.961–1.000 n = 25	r = 0.994 –1.000 n = 25
Initial reliability Asia Pacific		ICC = 0.977 95% CI: 0.910–0.999 n = 12	r = 0.979 –1.000 n = 12
Initial reliability Europe		ICC = 0.966 95% CI: 0.877–0.999 n = 17	r = 0.895 –1.000 n = 17
Retraining North America	ICC = 0.985 95%CI:0.928–1.000 n = 5	ICC = 0.973 95% CI: 0.901–0.999 n = 22	r = 0.988 –1.000 n = 22
Retraining Australia		ICC = 0.982 95% CI: 0.920–1.000 n = 6	r = 0.994 –1.000 n = 6
Retraining Japan/Asia Pacific		ICC = 0.979 95% CI: 0.914–0.999 n = 8	r = 0.986 – 0.999 n = 8
Retraining Europe		ICC = 0.982 95% CI: 0.930–1.000 n = 14	r = 0.971 –1.000 n = 14

ICC = Intraclass correlation coefficient; r = Pearson correlation coefficient.

The majority of evaluators were very experienced with the outcome measures and this was a crucial asset. Our training plan allowed for identification of those with less experience so that additional support could be provided. The fact that the measures were developed by an international group, many of who participated as sites for the trial provided a core group of experienced evaluators that had used the measures before and were comfortable with the population. Nevertheless, training for an international trial comes with challenges, including language and cultural differences between evaluators. The initial reliability results from the investigator meetings were used to adapt our training methodology as the trials progressed to allow for smaller groups and provide more support for language differences. Site visits were also provided with interpreters during training to assure reliability.

The initial training for the infantile onset trial was conducted in North America and Europe. The European evaluators were grouped with the Asia-Pacific evaluators in the European training. Even though translation was provided, the number of languages being translated created a learning experience that was less than optimal. This may have impacted the learning experience of the Asia-Pacific group and may have contributed to the somewhat lower

CHOP INTEND reliability in this group. The follow up annual training sessions for infantile onset trial and the initial and follow up training for the later onset SMA studies were further subdivided. The Asia-Pacific group was split into two groups with evaluators being trained separately and improved translation services were provided along with a higher ratio of trainers to evaluators to allow for more focused training and more individual attention. As a result the reliability in these groups improved and was similar across regions with this method of training.

Hands on assessment of reliability is not feasible because of the large number of evaluators across 3 large geographical regions who were involved in the trial. Repeated testing in infants and children is also limited due to behavior and fatigue. Video assessments are more practical but are limited by fixed visual field and lack of hands on contact but in this situation it was the best method. In this case the video inter-rater reliability we observed during training was also supported with administration of baseline and screening comparisons within the trial. This provided intra-rater reliability that was similar and reinforced the inter-rater reliability seen by video assessment. The comparison of screening and baseline evaluations allowed a combined assessment of both subject and administration variability within

one evaluator and the excellent intra-rater reliability supports the reliability of performance of evaluator, subjects, and the testing instruments within the trials.

We have shown that a rigorous training methodology can be deployed in the context of a global clinical trial. Disease specific measures can be administered reliably by a diverse group of evaluators with varied languages and cultures in infants and children with SMA. Here, we have compared within trial screening and baseline visits to assess intra-rater reliability and video based reliability assessment following the assessment training to monitor inter-rater reliability during the trial. By providing one on one support and periodic retraining to a group of therapist already with significant expertise in SMA, outcome reliability was able to be optimized. The commitment of the therapist evaluators to the training process and expertise in motor function along with their willingness to reach out to their one on one support trainers contributed to the success and reliability of outcome administration in these trials. In future trials, formal training and periodic retraining with reliability assessment can act to support evaluator consistency throughout the trial.

ACKNOWLEDGMENTS

The authors thank the patients and their families who participated in these studies as well as the clinical monitors, study coordinators, and evaluators. These studies were supported by Biogen and Ionis Pharmaceuticals, Inc.

DISCLOSURES

Allan Glanzman has received personal compensation for activities with Biogen as an advisory board member.

Elena Mazzone has received personal compensation for activities with Biogen, Ionis Pharmaceuticals, and Roche as a consultant.

Sally Dunaway Young has received personal compensation for activities with Ionis Pharmaceuticals as a consultant.

Richard Gee has been employed as a consultant Physical therapy trainer for Ionis pharmaceutical, has received personal compensation for activities with Ionis Pharmaceutical and Biogen as a consultant and has received research support from Ionis Pharmaceutical and Biogen.

Kristy Cokayne Rose has received personal compensation for activities with Ionis pharmaceutical as a consultant.

Anna Mayhew has received personal compensation for activities with Ionis Pharmaceutical as a consultant.

Leslie Nelson has received personal compensation for activities with Biogen, Ionis Pharmaceuticals, AveXis and Roche as a consultant.

Chris Yun has received personal compensation for activities with Ionis Pharmaceuticals, Emperra GMBH, Enzychem Lifesciences, Cardiora Pty Ltd. as a consultant.

Katie Alexander has received personal compensation for activities with Ionis Pharmaceuticals as an employee.

Basil T. Darras has been a scientific advisory board consultant for AveXis, Biogen, Cytokinetics, Marathon, PTC, Roche, and Sarepta; advisor for Ionis Pharmaceuticals, Inc.; research support from the National Institutes of Health/National Institute of Neurological Disorders and Stroke, the Slaney Family Fund for SMA, and the SMA Foundation; grants from Ionis Pharmaceuticals, Inc., Biogen, Cytokinetics, Fibrogen, PTC, Sarepta, and Summit.

Zarazuela Zolkipli has no conflicts to disclose.

Gihan Tennekoon has received personal compensation for activities with Biogen, Avexis, and PTC Therapeutics as an advisory board member.

John Day has received personal compensation for activities with Sarepta Therapeutics and PTC Therapeutics as a consultant.

Richard Finkel has received personal compensation for activities with Ionis Pharmaceuticals, Biogen, AveXis, Capricor, Catabasis, Lilly, Roche, Novartis; and the SMA Foundation, SMA Europe and Cure SMA as a consultant or advisor. Dr. Finkel has received research support from Ionis Pharmaceuticals, Biogen, Lilly, Cytokinetics Sarepta, NIH, MDA, and Summit

Eugenio Mercuri has received personal compensation for activities with Ionis Pharmaceuticals, Biogen, AveXis, Roche as a consultant or advisor.

Darryl De Vivo has received personal compensation for activities with AveXis, Biogen, Roche, IONIS, Sarepta, Cytokinetics Pharmaceuticals, and the SMA Foundation and has received research support from NIH, DOD, SMA Foundation, and Hope for Children Research Foundation.

Ron Baldwin Ron Baldwin has received personal compensation for activities with Ionis Pharmaceuticals as an employee.

Kathie Bishop received personal compensation for activities with Ionis Pharmaceuticals as an employee from 2009–2015. Dr. Bishop is a paid consultant to the SMA Foundation.

Jacqueline Montes receives support from the Eunice Kennedy Shriver National Institute for Child Health and Human Development (NICHD) 1K01HD084690-01A1 and reports serving as a consultant for Ionis. Pharmaceuticals, and Biogen, and serves on advisory boards for Roche Pharmaceuticals and Biogen.

REFERENCES

- [1] Finkel RS, McDermott MP, Kaufmann P, Darras BT, Chung WK, Sproule DM, et al. Observational study of spinal muscular atrophy type I and implications for clinical trials. *Neurology*. 2014;83(9):810-7.
- [2] Kaufmann P, McDermott MP, Darras BT, Finkel RS, Sproule DM, Kang PB, et al. Prospective cohort study of spinal muscular atrophy types 2 and 3. *Neurology*. 2012;79(18):1889-97.
- [3] Darras BT. Spinal muscular atrophies. *Pediatr Clin North Am*. 2015;62(3):743-66.
- [4] Glanzman AM, O'Hagen JM, McDermott MP, Martens WB, Flickinger J, Riley S, et al. Validation of the Expanded Hammersmith Functional Motor Scale in spinal muscular atrophy type II and III. *J Child Neurol*. 2011;26(12):1499-507.
- [5] Mazzone ES, Mayhew A, Montes J, Ramsey D, Fanelli L, Dunaway Young S, et al. Revised upper limb module for spinal muscular atrophy: Development of a new module. *Muscle Nerve*. 2016.
- [6] Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res*. 2005;19(1):231-40.
- [7] Rigo F, Hua Y, Krainer AR, Bennett CF. Antisense-based therapy for the treatment of spinal muscular atrophy. *J Cell Biol*. 2012;199(1):21-5.
- [8] Finkel RS, Mercuri E, Darras BT, Connolly AM, Kuntz NL, Kirschner J, et al. Nusinersen versus Sham Control in Infantile-Onset Spinal Muscular Atrophy. *N Engl J Med*. 2017;377(18):1723-32.
- [9] Chiriboga CA, Swoboda KJ, Darras BT, Iannaccone ST, Montes J, De Vivo DC, et al. Results from a phase 1 study of nusinersen (ISIS-SMN(Rx)) in children with spinal muscular atrophy. *Neurology*. 2016;86(10):890-7.
- [10] https://www.spinraza.com/content/dam/commercial/specialty/spinraza/caregiver/en_us/pdf/spinraza-prescribing-information.pdf [cited 2018 April 12].
- [11] <https://clinicaltrials.gov/ct2/show/NCT02292537?term=spinal+muscular+atrophy+nusinersen&rank=7> Accessed: May 19, 2017.
- [12] <https://clinicaltrials.gov/ct2/show/NCT02193074?term=spinal+muscular+atrophy+nusinersen&rank=11> Accessed: May 19, 2017.
- [13] Finkel R, Chiriboga C, Vajsar J, Day J, Montes J, De Vivo D, et al. Interim Results of a Phase 2 Clinical Study of Nusinersen (ISIS-SMNRx) in Patients with Infantile-Onset Spinal Muscular Atrophy. *Annals Of Neurology*. 2016;80(20):S370-S1.
- [14] Glanzman AM, Mazzone E, Main M, Pelliccioni M, Wood J, Swoboda KJ, et al. The Children's Hospital of Philadelphia Infant Test of Neuromuscular Disorders (CHOP INTEND): Test development and reliability. *Neuromuscul Disord*. 2010;20(3):155-61.
- [15] O'Hagen JM, Glanzman AM, McDermott MP, Ryan PA, Flickinger J, Quigley J, et al. An expanded version of the Hammersmith Functional Motor Scale for SMA II and III patients. *Neuromuscul Disord*. 2007;17(9-10):693-7.