Published in final edited form as:

J Pediatr. 2012 December ; 161(6): 1041–1046.e2. doi:10.1016/j.jpeds.2012.05.048.

Improving the Neonatal Research Network Annual Certification for Neurologic Examination of the 18–22 month Child

Jamie E. Newman, PhD^a, Carla M. Bann, PhD^a, Betty R. Vohr, MD^b, Anna M. Dusick, MD^c, and Rosemary D. Higgins, MD^d on behalf of the Follow-up Study Group of the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development Neonatal Research Network^{*}

^aStatistics and Epidemiology, RTI International, Research Triangle Park, NC

^bDepartment of Pediatrics, Women & Infants' Hospital, Brown University, Providence, RI

^cDepartment of Pediatrics, Indiana University School of Medicine, Indianapolis, IN

^d*Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD

Abstract

Objective—To describe the Neonatal Research Network's (NRN) efforts to improve the certification process for the Follow-up Study neurologic exam and to evaluate inter-rater agreement before and after two annual training workshops.

Study design—The NRN Follow-up Study is a multi-center observational study that has examined more than 11,500 infants from 1998–2010 and born 26 weeks gestational age at 18–22 months corrected age for neurodevelopmental outcome. The percentages of examiners who agreed with the Gold Standard examiner on four neurodevelopmental outcomes on the initial training video and a test video were calculated. Consistency among examiners was assessed with the first-order agreement coefficient (AC1) statistic.

Results—Improvements in agreement among examiners occurred between 2009 and 2010 and between initial training and test. Examiner agreement with the Gold Standard during the initial training was 83%–91% in 2009 and 89%–99% in 2010. Examiner agreement on the workshop test video increased from 2009 to 2010 with agreement reaching 100% for all four neurodevelopmental outcomes examined in 2010. AC1 values for the four neurodevelopmental outcomes on the training videos ranged from 0.64–0.82 in 2009 and 0.77–0.97 in 2010.

Conclusions—We demonstrate the importance of annual certification and the benefits of evaluation and revision of certification protocols to achieve high levels of confidence in neurodevelopmental study outcomes for multi-center networks.

© 2012 Mosby, Inc. All rights reserved.

Corresponding author: Jamie E. Newman, PhD, RTI International, 3040 Cornwallis Road, Research Triangle Park, NC 27709-2194, Phone: (919) 485-5719, Fax: (919) 485-7762, newman@rti.org.

^{*}A list of members of the Follow-up Study Group of the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development Neonatal Research Network is available at www.jpeds.com (Appendix).

The authors declare no conflicts of interest.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

examiner training; neurodevelopmental outcome; inter-rater agreement

Multi-center trials encounter challenges of standardizing research procedures across sites. It is important to periodically examine center differences, particularly for primary outcomes, to assess whether research procedures are being implemented uniformly. Periodical review of examiner training procedures can aid in identifying areas for improvement.

Examiner training for research protocol purposes is often achieved using video vignettes and to a lesser extent, the examination of actual patients during centralized training. Examiner training generally improves scoring accuracy and inter-rater reliability. For example, examiner training improved the motor development classification of infants as normal or abnormal¹ as well as the reliability of upper limb function scores.² However, there are instances where training has not improved inter-rater reliability and accuracy.³ Consideration has also been given to whether training and certification procedures are valid and reliable for novice examiners in addition to expert examiners.⁴

Training multiple examiners in a network for inter-rater reliability for the diagnosis of cerebral palsy (CP) and neurologic status can be particularly challenging due to substantial heterogeneity of neurologic findings ranging from very mild to severe, which is further complicated because young children may not clearly fall into a specific category. Examining the research standardization and training procedures of the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD) Neonatal Research Network (NRN) Follow-up Study may be beneficial for those conducting neurodevelopmental outcomes research in neonatal medicine as well as other fields where annual examiner certification is required. The NRN Follow-up Study evaluates infants that are born less than or equal to 26 completed weeks gestational age at 18 – 22 months corrected age for neurodevelopmental outcome. The Follow-up Study assessment battery includes: demographic and medical history, physical examination, a standardized neurologic examination,⁵ the Palisano et al^{6,7} Gross Motor Function Classification, a developmental assessment using the Bayley Scales of Infant and Toddler Development, 3rd edition (Bayley Scales-III),⁸ and a behavior assessment using the Brief Infant-Toddler Social Emotional Assessment (BITSEA).⁹ From 1998 to 2010, more than 11,500 infants completed the 18 – 22 month Follow-up Study visit.

Our objective is to describe the NRN's implementation of a protocol to improve the certification process for the neurologic exam by NRN center examiners by evaluating inter-rater agreement before and after the annual training workshops in 2009 and 2010. It was hypothesized that inter-rater agreement would improve between initial training and test as well as between 2009 and 2010.

Methods

The NICHD Neonatal Research Network maintains a study manual with standard definitions (Table I). A diagnosis of neurodevelopmental impairment (NDI) is implemented for both randomized trials and observational follow-up studies. The definition of NDI is the presence of any of the following: moderate to severe CP, bilateral blindness, bilateral hearing impairment requiring amplification, a gross motor function classification score (GMFCS)^{6,7} greater than or equal to level 2 reflecting moderate to severe motor impairment or a Bayley Scales-III cognitive or motor score less than 70. The NRN Follow-up Study neurologic examination⁵ includes a standard assessment of reflexes, muscle tone and strength,

functional motor skills, hearing and vision. Assessment vision and hearing is completed by: (1) reviewing history with family, (2) review of records from subspecialty ophthalmology and audiology services, (3) determining if child is receiving services for the blind, deaf educator, speech/language or sign language, or has amplification/FM system, (4) examination of the eyes including tracking, strabismus, glasses, surgical procedures, etc. and, (5) determining if the child can follow a simple verbal direction provided by the examiner. Hospital records, for the most part, contain this information and parents are usually aware of the formal diagnosis of blindness or permanent hearing loss.

Based on the neurologic examination, a determination is made of whether the child is normal or abnormal, and whether the child has CP as well as the severity of CP. CP is defined with the following three criteria: (1) Definite abnormalities observed in the classical neuromotor exam, which includes measurement of tone, deep tendon reflexes, coordination and movement. Any one definite abnormality in the classic neuromotor exam, as defined, except for isolated low tone (hypotonia) or toe walking without tight ankles is sufficient. (2) A delay in motor milestones with a disorder of motor function must be present. This may or may not be reflected in a motor quotient less than 70. In mild cases, there may be a subtle difference in hand functioning with a fine pincer grasp in one hand and a raking grasp in the other hand. Some disorder of motor function must be present. (3) Aberrations in primitive reflexes and postural reactions may be present.

The hierarchical classification of CP subtypes¹⁰ is included in the annual examiner training, with the aim of achieving inter-rater reliability in all subtypes including spastic, dyskinetic, and hypotonic CP. An assessment of gross motor function level is completed based on the Palisano et al.^{6,7} Gross Motor Function Classification system to provide a standardized classification of the severity of motor disability. This method is useful for assessing a level of severity of CP.^{6,7,11} Bimanual hand function is assessed in two ways. First, in a 3-level system that includes: (1) no problems with bimanual task, (2) some difficulty with bimanual tasks, and (3) no functional bimanual tasks, and second, in a 5-level assessment of each hand that ranges from 1 fine pincer grasp to 5 unable to grasp/makes no attempt to grasp.

Examiners

Since 2005, there have been 16 participating NRN centers. Each center designates a primary neurologic examiner, who is generally the Follow-up Study Principal Investigator (PI) of the center, based on their expertise in the field of newborn follow-up, neurologic examination of children, and developmental assessment. The primary neurologic examiner is responsible for fulfilling annual certification requirements and for training and certifying other examiners at their center to conduct the neurologic exam.

Certification Process

The annual certification process for the NRN has evolved over time (Table II; available at www.jpeds.com). The 1994 workshop protocol for the 1993 cohort was in place for five years (1994 to 1999). The training consisted of bringing six to seven children to the workshop at 18–22 months corrected age with a variety of neurologic diagnoses. Initially, there were 12 examiners who split into groups of three and rotated into the exam rooms to evaluate the children and score the neurologic exam sheets. The drawback with this approach was that by the time the third or fourth group entered the exam room, the child was intolerant of further examiners. Although this method of certification provided an opportunity to discuss the various diagnoses of each child, it was considered sub-optimal due to patient fatigue. In addition, not every examiner had a “hands-on” opportunity.

In 2000, the training protocol was modified to use videos of neurologic exams. Each center submitted a video of their site's primary neurologic examiner completing the neurologic examination on an 18–22 month corrected age former extremely low birth weight infant. The videos were all reviewed by two NRN Gold-Standard examiners and feedback was provided to each center examiner. Six high quality exams with a variety of diagnoses were identified for the training workshop. These six exams were viewed by the center examiners during the workshop, scored independently on the neurologic exam forms and discussed. Examiners reached consensus on the diagnoses after discussion and were certified as their center's primary neurologic examiner. An annual certification DVD was made with the selected exams following the workshop for the certified center examiners to certify additional developmentalists at their centers. During the 2008 workshop, however, it was proposed that the certification process be formalized further to train and clearly document inter-rater agreement on all test items including the NRN primary outcome of NDI, and to provide focused training during the workshops on areas of identified examiner weakness. Certification, however, would be determined by inter-rater agreement on four specific outcomes: normal neurologic exam, CP, GMFCS greater than or equal to level 2, and NDI.

Starting in 2009, although the initial submission process of videos by the primary examiners to the NRN Gold-standard examiners remained the same, the scoring of certification videos by each of the centers was formalized (Table II). Five to six selected exam videos that were representative of a spectrum of neurologic findings were collated on a certification DVD and sent to primary examiners to view and score using the NRN neurologic exam forms prior to the workshop. The forms were then keyed into the center's NRN data management system and transmitted to the data coordinating center at RTI International for analysis. The data from each center were then analyzed by the center examiner's correct response and by the group's consensus with the video findings. The NRN Gold-Standard examiners were then able to identify strengths and weaknesses of the center examiners' performance. With this process, the annual certification workshop in October 2009 began to focus training on scoring discrepancies and inter-rater agreement allowing for targeted training for problematic items. The workshop concluded with a "test" video, which examiners had not viewed before the annual certification workshop. Examiners viewed and scored the "test" video using NRN neurologic exam forms which were then entered into the NRN database for analysis. Examiners received feedback on both their training DVD and the scoring of their "test" video.

Data analyses

All analyses were performed using SAS 9.1 for Windows.¹² Statistical analyses focused on four primary outcomes of the neurologic assessment: (1) whether the neurologic exam was considered normal, (2) a diagnosis of moderate to severe CP, (3) a GMFCS greater than or equal to level 2 indicating moderate to severe gross motor impairment, and (4) whether the child met the neurologic or neurosensory component of the NRN definition of neurodevelopmental impairment (moderate/severe CP or GMFCS greater than or equal to level 2 or bilateral vision or hearing impairment). The percentage of examiners who agreed with the NRN Gold Standard raters on each video was computed. Consistency among examiners (i.e., inter-rater reliability) was assessed by computing the first-order agreement coefficient (AC1) statistic, using the AC1AC2 SAS macro.¹³ Even though the kappa statistic has traditionally been used to assess inter-rater reliability, researchers have noted several limitations of kappa, including the lack of correction for chance agreement and the kappa paradox wherein kappa can be low even though agreement is high.¹⁴ The AC1 statistic has been developed to address these limitations.¹⁵ To help with interpretation, we used the general guidelines for reliability coefficients suggested by Fleiss¹⁶ with values less than 0.40

indicating poor agreement, 0.40–0.75 indicating good agreement, and greater than 0.75 indicating excellent agreement.

Results

The four neurologic outcomes on the training videos for 2009 and 2010 are shown in Table III to illustrate the diagnoses chosen for the training process. In 2009, the centers' pre-workshop DVD scoring agreed with the Gold Standard rater 83% for normal neurologic exam to 91% for the GMFCS level 2 across the six videos (Table IV). During the workshop for real-time certification training in 2009, agreement on the test video increased to 100% for the diagnoses of abnormal neurologic exam and presence of CP, and agreement was 65% for GMFCS level 2 and NDI on test video 7. The findings of agreement and inconsistency for the 2009 exams are described in greater detail in an effort to clarify some of the challenges encountered in achieving consensus. Video 1 was scored as normal by the Gold Standard examiners. Differences were related to examiner perceptions of possible mild hip laxity, tight ankles, or a hearing loss. After discussion and reviewing, consensus was that the child was normal. For Video 2, some examiners had scored hypertonicity rather than spastic diplegia. Because resistance to dorsiflexion of the ankles and intermittent toe walking were present, the diagnosis of mild spastic diplegia was accepted. For Video 3, all examiners correctly classified CP, GMFCS and NDI, however, only 47% identified mild hypotonia with joint laxity. The lack of agreement on Video 4 was secondary to a diagnosis of hypotonia versus hypotonic CP with ataxia/athetosis. Because of associated movement disorder, a diagnosis of CP was considered correct. On Video 5, there was 100% agreement except on GMFCS level by one examiner. For Video 6, there was debate regarding the GMFCS level relative to the child's sitting posture. He was able to maintain a stable sit in a W-position or in a chair but not on the exam table. Final consensus was reached that it was GMFCS level 2 in a child with moderate spastic diplegia. The test Video 7 also was a child with spastic diplegia. The lack of agreement for this child was because of scoring the child as a GMFCS level 2 versus 1. The inter-rater analyses showed that weaker inter-rater agreement was more frequently present for diagnoses of disorders characterized by hypotonia and mild neurologic disorders.

Agreement of the centers' pre-workshop DVD scores in 2010 was higher than initial training in 2009 with values ranging from 89% for GMFCS level 2 to 99% for normal neurologic exam across the five videos. During the certification training in 2010, agreement on the test video reached 100% for all four outcomes. In addition, we computed the percentage of agreement between motor delay as defined as not walking at 18 months and the classifications for moderate/severe CP, GMFCS level 2 or higher, and NDI. Percentages of agreement were high, ranging from 82% to 83%.

The AC1 values show a similar pattern for agreement among examiners themselves (Table V). During 2009, AC1 values for the initial training phase ranged from 0.64 for normal neurologic exam to 0.82 for the GMFCS level 2, indicating good agreement. The examiners were entirely consistent with respect to normal neurologic exam and CP on the 2009 workshop test video (AC1=1.00), but somewhat less consistent on the GMFCS level and NDI for this particular video (AC1=0.24).

The examiners demonstrated good to excellent agreement during the initial training in 2010 with AC1 values ranging from 0.77 for GMFCS level to 0.97 for normal neurologic exam. All examiners were consistent in their ratings of the workshop test video in 2010 as reflected in the values of 1 for AC1.

Discussion

Challenges remain in the attempt to achieve inter-rater reliability for the neurologic exam and for the diagnosis of CP in young children. Although use of the GMFCS in conjunction with the classic neurologic exam has improved examiner accuracy, diagnosing severity of impairment is not a perfect science.¹¹ A number of investigators have made substantial progress in initiatives to simplify and categorize the methodology for diagnosing CP.^{10,17,18} The NRN has begun to use the classification system developed by the Surveillance of Cerebral Palsy in Europe working group.¹⁰ The definition and classification of CP was recently extensively reviewed and updated in a supplement to *Developmental Pediatrics and Child Neurology*.¹⁸ Implementation of this new system and categorization will allow for greater agreement for each extremity involved. The revised system includes a flow chart which has definitions for specific clinical features applied to each extremity. The NRN has found that a number of former preterm children have mixed findings which has made it challenging to place these children in a specific category of CP based on the classic neurologic exam. Implementation of the revised system should produce enhanced reliability for networks. A second enhancement to be considered is the new Bimanual Fine Motor Function Test reported by Beckung and Hagberg¹⁹ which uses a 5-level classification system for the upper extremities similar to the GMFCS for the lower extremities.

Despite limitations, improvements were observed in examiner agreement with the new NRN neurologic exam certification process between 2009 and 2010. Examiner agreement with the NRN Gold Standard during the initial training in 2010 was higher than initial training in 2009. Also, an increase in examiner agreement on the workshop “test” video occurred between 2009 and 2010 with agreement reaching 100% for all four neurologic outcomes examined in 2010. The AC1 values provided further support for these improvements.

There are several benefits to the new certification process. First, quantifying the certification process facilitated evaluation of annual certification by making it possible to assess inter-rater agreement. Second, annual neurologic exam training is now focused on scoring discrepancies which allows for targeted training of problematic items. Third, examiners who are outliers can be identified using the standardized videos, and can receive additional training, thus improving cohort data collection in this difficult area of judgment. Fourth, face-to-face meeting time for training is shorter than it was before 2009 because examiners view the training videos and score them prior to the training workshop. The time previously spent on scoring exams during the training workshop is now allocated towards setting the NRN Follow-up Study research agenda and discussing new protocols.

A limitation of this certification process is the difficulty of seeing all angles clearly and inability to assess muscle tone by palpation on training videos. To address this, examiners were asked to narrate the neurologic exam as it is being video-recorded so that the viewer has knowledge of the tactile information that is needed to score the exam. The only child information included on the DVDs was the child’s age at the time of the exam. A complete medical history was included in the early years of the NRN though this information is no longer included due to the potential for examiner bias.

The study findings clearly demonstrate the importance of annual certification and assessment of inter-rater agreement as well as the benefits of evaluation and revision of certification protocols to achieve high levels of confidence in neurodevelopmental study outcomes for multi-center networks. The NRN will continue to review processes for training and annual certification to identify areas for improvement. Evaluating inter-rater agreement before and after annual training workshops is one way to identify areas where additional training may aid in the standardization of research procedures across sites.

Acknowledgments

The National Institutes of Health and the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD) provided grant support for the Neonatal Research Network's Follow-up Study. Data collected at participating sites of the NICHD Neonatal Research Network were transmitted to RTI International, the data coordinating center for the network, which stored, managed and analyzed the data for this study. As the Data Coordinating Center coordinator for the Follow-up Study, J.N.'s time was supported by the NICHD Neonatal Research Network.

We are indebted to our medical and nursing colleagues and the infants and their parents who agreed to take part in this study.

List of Abbreviations

NICHD	<i>Eunice Kennedy Shriver</i> National Institute of Child Health and Human Development
NRN	Neonatal Research Network
Bayley Scales-III	Bayley Scales of Infant and Toddler Development 3rd edition
BITSEA	Brief Infant-Toddler Social Emotional Assessment
NDI	neurodevelopmental impairment
CP	cerebral palsy
GMFCS	Gross Motor Function Classification Score

References

1. Blanchard Y, Neilan E, Busanich J, Garavuso L, Kilmas D. Interrater reliability of early intervention providers scoring the Alberta Infant Motor Scale. *Pediatr Phys Ther.* 2004; 16:13–18. [PubMed: 17057466]
2. Cusick A, Vasquez M, Knowles L, Wallen M. Effect of rater training on reliability of Melbourne Assessment of Unilateral Upper Limb Function scores. *Dev Med Child Neurol.* 2005; 47:39–45. [PubMed: 15686288]
3. Cook DA, Dupras DM, Beckman TJ, Thomas KG, Pankratz VS. Effect of rater training on reliability and accuracy of mini-CEX scores: a randomized, controlled trial. *J Gen Intern Med.* 2009; 24:74–79. [PubMed: 19002533]
4. Lyden P, Raman R, Liu L, Emr M, Warren M, Marler J. National Institutes of Health Stroke Scale certification is reliable across multiple venues. *Stroke.* 2009; 40:2507–2511. [PubMed: 19520998]
5. Amiel-Tison, C.; Gosselin, J. *Neurological Development from Birth to Six Years; Guide for Examination and Evaluation.* Johns Hopkins University Press; Baltimore, Maryland: 2001.
6. Palisano RJ, Hanna SE, Rosenbaum PL, Russell DJ, Walter SD, Wood EP, et al. Validation of a model of gross motor function for children with cerebral palsy. *Phys Ther.* 2000; 80:974–85. [PubMed: 11002433]
7. Palisano RJ, Rosenbaum PL, 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; 39:214–233. [PubMed: 9183258]
8. The Psychological Corporation. *Bayley Scales of Infant and Toddler Development.* 3. San Antonio, TX: Psychological Corp; 2006.
9. Briggs-Gowan MJ, Carter AS, Irwin JR, Wachtel K, Cicchetti DV. The Brief Infant-Toddler Social and Emotional Assessment: screening for social-emotional problems and delays in competence. *J Pediatr Psychol.* 2004; 29:143e55. [PubMed: 15096535]
10. Surveillance of Cerebral Palsy in Europe (SCPE). Surveillance of cerebral palsy in Europe: a collaboration of cerebral palsy surveys and registers. *Dev Med Child Neurol.* 2000; 42:816–824. [PubMed: 11132255]

11. Reid SM, Carlin JB, Reddihough DS. Using the Gross Motor Function Classification System to describe patterns of motor severity in cerebral palsy. *Dev Med Child Neurol.* 2011; 53:1007–1012. [PubMed: 22014320]
12. The SAS Institute Inc. The SAS System for Windows. Ver 9.1. Cary, NC: The SAS Institute Inc; 2002–2003.
13. Blood, E.; Spratt, KF. [Accessed March 4, 2011] Disagreement on Agreement: Two Alternative Agreement Coefficients. SAS Global Forum. 2007. Paper 186–2007. Available at: <http://www2.sas.com/proceedings/forum2007/186-2007.pdf>
14. Uebersax, J. [Accessed March 4, 2011] Kappa Coefficients. 2002. Available at: <http://www.john-uebersax.com/stat/kappa.htm>
15. Gwet, KL. Handbook of Inter-Rater Reliability. 2. Gaithersburg, MD: Advanced Analytics, LLC; 2010.
16. Fleiss, JL. Statistical Methods for Rates and Proportions. 2. New York: Wiley; 1981.
17. Evans P, Johnson A, Mutch L, Alberman E. A standard form for recording clinical findings in children with a motor deficit of central origin. *Dev Med Child Neuro.* 1989; 31:119–127.
18. Definition and Classification of Cerebral Palsy. *Dev Med Child Neurol.* 2007; 49(Supplement109): 1–44. [PubMed: 17371509]
19. Beckung E, Hagberg G. Neuroimpairments, activity limitations and participation restrictions in children with cerebral palsy. *Dev Med Child Neurol.* 2002; 44:309–316. [PubMed: 12033716]

Table 1

NICHD Neonatal Research Network 2009–2010 Definitions for the 18–22 Month Neurologic Exam

Parameter	Definition
Normal	No abnormalities of tone, reflexes or function noted in the neurologic exam.
Cerebral Palsy	Abnormalities observed in tone, deep tendon reflexes, coordination and movement. Some disorder of motor function must be present Aberrations in primitive reflexes and postural reactions may be present.
GMFCS	Normal - Walks 10 steps independently and fluently Possible Level 1 - Walks 10 steps independently but not fluently; child exhibits toe walking (mild diplegia) or asymmetric walking (possible hemiplegia) Level 1 - Infants move in and out of sitting and floor-sit with both hands free to manipulate objects. Infants creep or crawl on hands and knees, pull to stand and take steps holding onto furniture. Infants walk between 18 months and 2 years of age without holding on Level 2 - Infants maintain floor sitting but may need to use their hands for support to maintain balance. Infants creep on their stomach or crawl on hands and knees. Infants may pull to stand and take steps holding onto furniture Level 3 - Infants maintain floor sitting when the low back is supported. Infants roll and creep forward on their stomachs Level 4 - Infants have head control but trunk support is required for floor sitting. Infants can roll to supine and may roll to prone Level 5 - Physical impairments limit voluntary control of movements. Infants are unable to maintain antigravity head and trunk postures in prone and sitting. Infants require adult assistance to roll.
Neurodevelopmental Impairment	The presence of any one of the following: Moderate to Severe CP with a GMFCS score of 2, bilateral blind with vision 20/200 or worse, bilateral permanent hearing loss $\pm$ amplification that does not permit the child to understand instructions of the examiner, or a Bayley Scales-III cognitive or motor score < 70.

Table 2

Methodology for Neurologic Exam Training on Former-Preterm Children at a Corrected Age of 18–22 Months between 1994 and 2010

Time Period	1994–1999	2000–2008	2009–2010
Process	6–8 children with a spectrum of diagnoses brought in for assessment.	Each center's neurologic examiner submitted a video of self examining a child to 2 NRN Gold Standard examiners and feedback was provided.	Each center's neurologic examiner submitted a video of self examining a child. The 2 NRN Gold Standard examiners scored each video and achieved consensus on the 16 DVDs and feedback was provided.
Methods	Children examined by groups of 3–4 neurologic examiners and forms scored. Often 1 examiner did the neurologic exam. All examiners completed forms.	6–7 videos with a variety of diagnoses were chosen for training at the annual workshop. Forms were completed at the workshop while viewing videos.	6–7 DVDs with a variety of diagnoses were sent to centers to be viewed, scored and entered into the NRN database prior to the workshop. Inter-rater agreement to the NRN Gold Standard Examiners was calculated.
Certification	Groups met and discussed diagnoses on forms and came to consensus.	The training videos and diagnoses on forms discussed and consensus reached at the workshop.	The workshop focused on areas with weak inter-rater agreement and a final “test” video was viewed and scored during the workshop and entered into the NRN database for analysis and certification.

Table 3

Neurologic Diagnoses for 2009 and 2010

2009 Videos	Normal	Cerebral Palsy	GMFCS 2	Neurodevelopmental Impairment (NDI)
1	yes	No	No	No
2	No	CP: Spastic diplegia	No	No
3	No	Generalized hypotonia	No	No
4	No	CP: hypotonic ataxic	Yes	Yes
5	No	CP: Spastic quadriplegia	Yes	Yes
6	No	CP: Spastic diplegia	Yes	Yes
Test Video 7	No	CP: Spastic diplegia	Yes	Yes
2010 Videos	Normal	Cerebral Palsy	GMFCS 2	Neurodevelopmental Impairment (NDI)
1	No	Generalized hypotonia	No	No
2	Yes	No	No	No
3	No	CP: varying tone & athetosis; Choreo-athetotic CP	Yes	Yes
4	No	CP: hypotonic ataxic	Yes	Yes
5	No	CP: Spastic quadriplegia	Yes	Yes
Test Video 6	No	CP: Spastic triplegia	Yes	Yes

Table 4
Percentage of Examiners Agreeing with Gold Standard Rater by Training Session and Video for 2009 and 2010

Training Session Video		Outcome				
	Diagnosis	Normal Neurological Exam	Cerebral Palsy	GMFCS Level	GMFCS Level 2	NDI
2009						
Pre workshop score						
Video 1	Normal	53	100	89	100	74
Video 2	CP-Spastic diplegia	100	63	92	100	100
Video 3	Hypotonia	47	100	84	100	100
Video 4	CP-Athetosis/dystonia	95	74	57	100	100
Video 5	CP-Spastic quadriplegia	100	100	63	95	100
Video 6	CP-Spastic diplegia	100	89	53	53	58
Videos 1-6		83	88	73	91	89
Workshop Test						
Video 7	CP-Spastic diplegia	100	100	65	65	65
2010						
Pre workshop score						
Video 1	CP-Generalized hypotonia	100	75	92	92	94
Video 2	Normal	94	100	100	100	100
Video 3	CP-Athetosis	100	100	81	100	100
Video 4	CP-Hypotonia with ataxia	100	100	50	50	63
Video 5	CP-Spastic quadriplegia	100	100	94	100	100
Videos 1-5		99	95	83	88	91
Workshop Test						
Video 6	Spastic triplegia	100	100	100	100	100

Table 5
AC1 Measures* of Agreement among Examiners by Training Session and Video for 2009 and 2010

Training Session Video	Diagnosis	Outcome				
		Normal Neurological Exam	Cerebral Palsy	GMFCS Level	GMFCS Level 2	NDI
2009						
Pre workshop score						
Video 1	Normal	0.02	1.00	0.77	1.00	0.42
Video 2	CP-Spastic diplegia	1.00	0.20	0.82	1.00	1.00
Video 3	Hypotonia	-0.05	1.00	0.66	1.00	1.00
Video 4	CP-Athetosis/dystonia	0.89	0.42	0.08	1.00	1.00
Video 5	CP-Spastic quadriplegia	1.00	1.00	0.20	0.89	1.00
Video 6	CP-Spastic diplegia	1.00	0.77	0.02	0.02	0.10
Videos 1–6		0.64	0.73	0.43	0.82	0.75
Workshop test						
Video 7	CP-Spastic diplegia	1.00	1.00	0.24	0.24	0.24
2010						
Pre workshop score						
Video 1	CP-Generalized hypotonia	1.00	0.45	0.82	0.82	0.87
Video 2	Normal	0.87	1.00	1.00	1.00	1.00
Video 3	CP-Athetosis	1.00	1.00	0.59	1.00	1.00
Video 4	CP-Hypotonia with ataxia	1.00	1.00	-0.02	-0.02	0.18
Video 5	CP-Spastic quadriplegia	1.00	1.00	0.87	1.00	1.00
Videos 1–5		0.97	0.89	0.65	0.76	0.81
Workshop test						
Video 6	Spastic triplegia	1.00	1.00	1.00	1.00	1.00

* Values less than 0.40 indicate poor agreement, 0.40–0.75 good agreement, and greater than 0.75 excellent agreement.