# INTERRATER RELIABILITY OF ELECTRODIAGNOSIS IN NEONATAL BRACHIAL PLEXOPATHY

**Mary Catherine Spires, MD[1], Spencer M. Brown, MS[3], Kate Wan-Chu Chang, MA, MS[2], James A. Leonard, MD[1], Lynda J-S. Yang, MD, PhD[2]**

Departments of [1]Physical Medicine and Rehabilitation and [2]Neurosurgery, University of Michigan, Ann Arbor, Michigan, USA

[3]Tulane University, New Orleans, LA, USA

**CORRESPONDENCE:**

Mary Catherine Spires, MD

Department of Physical Medicine and Rehabilitation

University of Michigan

325 East Eisenhower

Ann Arbor, MI 48108

Phone: 734 764 8347; Fax: 734 615 6713

Email: mcspires@med.umich.edu

**RUNNING TITLE:** EDX on NBPP

**FINANCIAL DISCLOSURES:**

No funding was received in relationship to this manuscript.

**CONFLICT OF INTEREST:**

The authors have no conflicts of interest to report pertaining to the materials or methods used in this study or the findings specified in this paper.

## ABSTRACT

***Introduction:*** We investigated interrater reliability of overall assessment of nerve root lesions by electrodiagnostic testing (EDX) in neonatal brachial plexus palsy (NBPP).

***Methods:*** Two blinded, board-certified reviewers retrospectively reviewed de-identified EDX data from 37 infants with NBPP for 2005-2012. Only nerve conduction and electromyography needle data were included. The examiners independently assigned 1 of 4 nerve root lesion categories: 1) normal, pre-ganglionic lesion (avulsion), 2) post-ganglionic lesion (rupture), 3) normal, or 4) "unable to determine." Simple percentage agreement, the Cohen kappa statistic representing interrater reliability for each nerve root (C5-T1), and overall kappa between examiners were evaluated.

***Results:*** Interrater reliabilities were substantial to almost perfect for each nerve root except C5. Considering all nerve roots, overall interrater reliability was substantial (kappa=0.62); simple percentage agreement was 75% (138/185).

***Conclusions:*** Interrater reliability of nerve root assessment by EDX for infants with NBPP was high for C6-T1 root levels, but less reliable for C5 because of technical factors.

**KEY WORDS:** Brachial plexopathy; electrodiagnosis; electromyography; EMG; infants; interrater reliability; neonates; nerve conduction; pediatrics

## INTRODUCTION

Neonatal brachial plexus palsy (NBPP) is a relatively common problem with an incidence ranging from 0.4% to 4% of live births.[1-3] While the majority of these plexopathies resolve spontaneously, a cohort exists that has persistent long-term deficits. Assessing the anatomic level and severity of nerve root lesions in NBPP is challenging. Distinguishing pre-ganglionic lesions from post-ganglionic lesions is fundamental for determining optimal medical and surgical management to maximize functional outcome.[4] Pre-ganglionic lesions, or root avulsions, are severe injuries that can extend into the transitional zone of the spinal cord. Spontaneous recovery does not occur, and prognosis is poor. No effective surgical repair of pre-ganglionic lesions is available. Post-ganglionic lesions require early and close follow-up, as the time frame for primary neurosurgical repair is limited to several months.

The gold standard for assessing a nerve root lesion is direct observation during surgical exploration of the brachial plexus; however, surgical exploration carries significant risk and is not always indicated or appropriate. Several diagnostic approaches are used to assess the anatomic location and severity of a brachial plexopathy.[5] Electrodiagnostic testing (EDX) is a noninvasive method employed commonly to assess lesion site, type of lesion, and extent of neural dysfunction. EDX consists of nerve conduction studies (NCS) and needle electromyography (EMG) of muscles to quantify the presence or absence of denervation as well as voluntary MUP activity and morphology.

EDX is performed in infants to characterize nerve root lesions in brachial plexopathy, but its use is controversial. EDX of infants is often criticized as subjective, and interpretation of results is inconsistent. Adding to the controversy is the fact that overall infant size and limb size

are smaller, making localization of specific muscles more challenging than in adults. Motor unit potential (MUP) amplitudes are smaller and often biphasic, rather than triphasic as in adults.[6,7] These different MUP characteristics in infants may mislead the inexperienced electromyographer. Additionally, infants do not move on request, and limb movements are not graded or controlled, thus assessment of MUP recruitment is more challenging. Often maximal MUP recruitment is observed at initial electrode insertion, adding further difficulty to interpretation of the examination.

The interrater reliability of assessment of nerve root lesions by EDX has not been studied in infants, particularly in the brachial plexus. In a review of the literature, 2 studies were identified that focused on interrater reliability; both studied adults. Chouteau et al.[8] and Kendall et al.[9] investigated interrater reliability of the electrodiagnostic impression in lumbar radiculopathy. No studies focusing on brachial plexopathy, adult or pediatric, were identified. Thus, the goal of our study was to investigate interrater reliability of the assessment of nerve root lesions by EMG and NCS performed on infants with NBPP.

## MATERIALS AND METHODS

We performed a retrospective review of 37 non-sedated infants with the clinical diagnosis of NBPP who underwent EDX. All subjects were referred to the NBPP multi-specialty clinic in a large academic center. The study period was 2005 to 2012. Criteria for study inclusion were: 1) full-term infants who were referred for evaluation of NBPP; 2) EDX was performed within 45 days of birth; 3) EDX was performed at 1 institution by an American Board of Electrodiagnostic Medicine (ABEM)-certified physician. Patients were excluded from study if they had additional

neurological or musculoskeletal disorders, congenital limb malformations, or hereditary disorders. A total of 37 infant EDX studies met the inclusion criteria. Data collected included age at initial appointment, age at EDX study, gender, and race. Detailed clinical examinations were performed on the affected arms of all subjects. The side of NBPP and Narakas severity grade[10-12] were also included in the data analysis. This study was approved by the University of Michigan Institutional Review Board; because patient data were de-identified, individual patient informed consent was not required.

**EDX and Blinded EDX Review**

EDX examination of the brachial plexus was performed by 4 independent examiners who were all certified by the ABEM. The examiners who performed the EDX obtained a history and physical examination for each infant prior to the study. NCS were performed on motor and sensory nerves of the involved upper extremity. Sensory studies were performed antidromically, while motor studies were performed orthodromically. Median and ulnar sensory nerve action potentials were tested, and the amplitude, distal latency, and conduction velocities were recorded for each infant. Median and ulnar compound motor action potentials were obtained, and distal latency, amplitude, and conduction velocities recorded for each infant.

EMG examination was performed using pediatric concentric needle electrodes on relevant muscle sets. Two blinded reviewers who practice at the same institution and are both certified by the ABEM reviewed de-identified EDX data. EDX data included only the NCS and EMG results without raw traces. No history or physical examination information was provided. NCS data included the nerves studied, response amplitude, distal latency, and conduction velocity. The EMG data provided to the reviewers included insertional activity, presence or

absence of spontaneous activity (positive waves, fibrillation potentials, and fasciculation potentials) as well as the MUP morphology. With the data provided by NCS and EMG, each reviewer was asked to give an assessment of each nerve root using 1 of the following 4 electrodiagnostic impressions: 1) pre-ganglionic (avulsion) lesion, 2) post-ganglionic (rupture) lesion, 3) normal, or 4) "unable to determine" if the lesion was pre- or post-ganglionic for the nerve roots tested.

**Statistical Analysis**

Patient demographics, NCS, and EMG data were reported with descriptive statistics. Interrater reliability refers to the agreement or concordance among reviewers. A score of homogeneity or consensus can be assigned. The Cohen kappa statistic was assessed between the 2 blinded reviewers to represent interrater reliability with regard to the 4 options of EDX impression.[13] Percentage agreement refers to the raw numerical value for the number of agreement of total assessment at each nerve root. Because the Cohen kappa analysis considers both the observed agreement and the expected agreement by chance, it is a more robust representation of interrater reliability than the simple percentage agreement calculation. The values of Kappa statistics range from 0 to 1: 0.01–0.20 indicates slight agreement, 0.21–0.40 fair agreement, 0.41–0.60 moderate agreement, 0.61–0.80 substantial agreement, and 0.81–0.99 almost perfect agreement.[13,14] To maintain an evaluation-wise error rate of 0.05, Bonferroni correction for 37 subjects with *P*-values < 0.001 were considered statistically significant. Statistical analyses were performed using SPSS software, version 21 (SPSS Inc., Chicago, IL).

## RESULTS

Data from the EDX of 37 infants were included in the analysis. Demographic and clinical information is provided in Table 1. Information on NCS is found in Table 2. All patients were tested only on the involved side, except for 1 patient who underwent bilateral upper extremity testing. The deltoid, biceps brachii, triceps, and first dorsal interosseous muscles were examined most frequently by EMG (Table 3).

Percentage agreement, kappa of each nerve root (C5-T1), and overall kappa between examiners are presented in Table 4. Table 5 represents percentage agreement and kappa score by upper (C5-6), middle (C7), and lower (C8-T1) trunks. Interrater reliability for each nerve root ranged from substantial to almost perfect, except for C5 (fair, kappa=0.27, percentage agreement=38%). At the nerve trunk level, the upper trunk had moderate (kappa=0.43, percentage agreement=58%) interrater reliability, while middle and lower trunks had substantial or better interrater reliability (kappa=0.86 and 0.70, respectively). Overall interrater reliability was substantial (kappa=0.62) with 75 (138 of 185) percentage agreement.

## DISCUSSION

Interrater reliability of the overall assessment of nerve root lesions by EDX in infants with NBPP has not been studied. Although there is literature to support interrater reliability of the individual components of electrodiagnosis (ie, NCS and EMG), this study takes the next step by using both NCS and EMG to evaluate interrater reliability of the lesion in each nerve root to determine whether the lesion is pre- or post-ganglionic.

The 2 studies identified in the literature that assessed interrater reliability were limited to adults referred for lower extremity evaluation. Chouteau et al.[8] reported a very high interrater reliability (kappa >0.90). Kendall et al.[9] reported the sensitivity of the EMG study to be 79% when it was used to identify a radiculopathy without specifying the level of the lumbar lesion. Both studies were limited to EMG only and did not include NCS.

The findings of our study support a high interrater reliability for assessment of nerve root lesions (pre- or post-ganglionic) by EDX when studies are interpreted by 2 blinded, ABEM-certified electrodiagnosticians. The lowest interrater reliability was for the C5 nerve root (kappa 0.27). Several possibilities may have led to this level of disagreement, but most likely it is based on the nerves and muscles that were chosen for testing by the electrodiagnosticians who performed the study. Currently no standard protocol or guideline exists or is agreed upon in the literature for electrodiagnostic evaluation of brachial plexopathy.

The studies reviewed contained limited electrodiagnostic data about the C5 nerve root, which likely contributed to the lower interrater reliability. Each reviewer may have used different criteria to determine whether a lesion was pre- or post-ganglionic, normal, or "unable to determine." Additionally, reviewers may have had different thresholds for what they consider sufficient data to localize the lesion and therefore called the lesion "unable to determine." There are several technical NCS and EMG issues that limit EDX assessment of the C5 nerve root in neonates. In the studies reviewed, no NCS were performed that were specific for the C5 nerve root. Median nerve sensory testing was performed using the index finger for recording, which provides information about C6 function. The index finger is cited as being innervated by C6 in sources.[15-18] It is noted that some authors also consider the index finger as being innervated by C6/7 or C7.[19] However, we must point out that the results from a single nerve evaluation alone

were not used to localize the lesion. The data were analyzed in conjunction with the results from other nerves studied and the findings on needle examination. Median sensory NCS performed with recording from the thumb of an infant was not done likely due the technical difficulties inherent in testing on such a small digit. NCS of the lateral antebrachial cutaneous nerve assesses the C5 root function. NCS of this nerve is very difficult technically, particularly in awake and moving infants, and was not done in these cases.

Further problems arise in distinguishing C5 versus C6 nerve function on EMG. The rhomboid major and minor muscles are considered to be the only predominately C5 innervated muscles. In infants, it is impractical to test these muscles due to their size, location under the trapezius muscle, and proximity to underlying thoracic contents. Since an infant cannot cooperate with the test, it is not possible to perform any maneuvers to demonstrate that the electromyographer had definitively isolated these muscles from the overlying trapezius. Despite these difficulties, Vanderhave et al.[4] reported that overall, EDX were most useful in identifying post-ganglionic lesions within the plexus, particularly in the upper plexus. The infraspinatus and supraspinatus muscles are also innervated by C5 and C6 and could potentially serve as a source of additional information about the function of C5 more proximally. However, scapular ossification is incomplete, increasing the risk of entry into the thorax and subsequent pneumothorax while trying to test the infraspinatus or supraspinatus muscles. The supraspinatus and infraspinatus muscles also lie under the trapezius, making isolation difficult, as noted above. Interrater reliability was the greatest for C7 (kappa 0.86). The triceps is easily accessed for needle examination due to its superficial anatomy, and it is commonly studied for C7 involvement, though some C8 innervation is present. It is a larger muscle than others in the upper extremity so sampling errors can be made, that is, the electromyographer may have sampled a

part that was not involved. It is unlikely that all fibers of the triceps muscle were equally affected. No data regarding the median sensory nerve response from the middle finger were available. Lastly, NBPP lesions are predominately upper trunk, and C7 abnormalities are less common. Since the incidence of C7 involvement is less common, the reviewers may have been less likely to judge this nerve root as abnormal.

It is worth noting that none of these infants were sedated for EDX, and they were awake and active. Sedation is used in some settings, but it is not without risk in this population. Some electrodiagnosticians perform the EDX after sedation for MRI.[20] However, not all infants undergo an MRI in this setting, unless neurosurgical intervention is being considered or scheduled.

**Study Limitations**

There are limitations to this study. First this was a retrospective review of data collected over a 7-year period (2005-2012). As the electrodiagnosticians gained experience, the thoroughness of the examination and expertise in performing these tests on infants may have improved. The 2 reviewers performed some but not all of the EDX studies. All of the studies reviewed were de-identified so that the reviewers could not identify the studies they performed personally. Additionally, the reviewers could only assess each nerve root lesion (pre- or post-ganglionic) from the NCS and EMG data they were provided. Unlike an electrodiagnostician doing the EDX, the reviewers did not have the ability determine the number of muscles to be evaluated, how thoroughly a muscle segment was studied (eg, number of quadrants examined), whether to consider using additional techniques such as triggering, to assess MUPs, or have the ability to repeat examination of a muscle if needed for clarification of the EMG findings to arrive at a

nerve root assessment. Interrater reliability should not be mistaken for reproducibility, since actual performance of the diagnostic procedure was not evaluated. Additionally, the data were evaluated by 2 blinded reviewers who practice at the same academic institution. The electrodiagnosticians who performed the studies were trained at various institutions. All electrodiagnosticians involved, including the reviewers, are ABEM-certified. Lastly, no electrodiagnostic standard protocol or guidelines for infants with brachial plexopathy exist in the literature or are recommended by the American Association of Neuromuscular Disease and Electrodiagnostic Medicine.

## CONCLUSIONS

The interrater reliability of each nerve root lesion assessment by EDX performed on infants with NBPP is substantial for the C6-T1 root levels when studies are performed by expert electrodiagnosticians. Interpretations related to the C5 root are less reliable. It is likely that technical factors and the limited NCS and EMG options for distinguishing C5 root function from the C6 root increases the difficulty of specifying the presence and/or extent of a C5 root lesion, thereby limiting adequate data to determine the function of the C5 nerve root.

## ABBREVIATIONS

ABEM = American Board of Electrodiagnostic Medicine

EDX = electrodiagnostic testing

EMG = electromyography

MUP = motor unit potential

NBPP = neonatal brachial plexus palsy

NCS = nerve conduction studies

## REFERENCES

1. Doumouchtsis SK, Arulkumaran S. Are all brachial plexus injuries caused by shoulder dystocia? Obstet Gynecol Surv 2009;64:615-623.
2. Gilbert WM, Nesbitt TS, Danielsen B. Associated factors in 1611 cases of brachial plexus injury. Obstet Gynecol 1999;93:536-540.
3. Piatt JH, Jr. Birth injuries of the brachial plexus. Clin Perinatol 2005;32:39-59.
4. Vanderhave KL, Bovid K, Alpert H, Chang KW, Quint DJ, Leonard JA, Jr., Yang LJ. Utility of electrodiagnostic testing and computed tomography myelography in the preoperative evaluation of neonatal brachial plexus palsy. J Neurosurg Pediatr 2012;9:283-289.
5. Somashekar D, Yang LJ, Ibrahim M, Parmar HA. High-resolution MRI evaluation of neonatal brachial plexus palsy: A promising alternative to traditional CT myelography. AJNR Am J Neuroradiol 2014;35:1209-1213.
6. Carmo R. Motor unit action potential parameters in human newborn infants. Arch Neurol 1960;3:136-140.
7. Sacco G, Buchthal F, Rosenfalck P. Motor unit potentials at different ages. Arch Neurol 1962;6:366-373.
8. Chouteau WL, Annaswamy TM, Bierner SM, Elliott AC, Figueroa I. Interrater reliability of needle electromyographic findings in lumbar radiculopathy. Am J Phys Med Rehabil 2010;89:561-569.
9. Kendall R, Werner RA. Interrater reliability of the needle examination in lumbosacral radiculopathy. Muscle Nerve 2006;34:238-241.

10. Gilbert A, Tassin JL. [Surgical repair of the brachial plexus in obstetric paralysis]. Chirurgie 1984;110:70-75.

11. Narakas AO. Injuries to the brachial plexus. In: Bora FW, editor. The Pediatric Upper Extremity: Diagnosis and Management. Philadelphia: Saunders; 1986. p 247-258.

12. Narakas AO. Obstetrical brachial plexus injuries. In: Lamb DW, editor. The Paralysed Hand. New York: Churchill Livingstone; 1987. p 116-135.

13. Cohen J. Coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20:10.

14. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-174.

15. Foerster O. The dermatomes in man. Brain 1933;56:1-39.

16. Grant JCB. Grant's Atlas of Anatomy. Baltimore: Williams & Wilkins; 1972.

17. Haymaker WE, Woodhall B. Peripheral Nerve Injuries: Principles of Diagnosis. Philadelphia: Saunders; 1953.

18. Kimura J. Electrodiagnosis in Diseases of Nerve and Muscle: Principles and Practice. New York: Oxford University Press; 2013.

19. Preston DC, Shapiro BE. Electromyography and Neuromuscular Disorders: Clinical-Electrophysiologic Correlations. New York: Saunders; 2013.

20. Ing C, DiMaggio C, Whitehouse A, Hegarty MK, Brady J, von Ungern-Sternberg BS, Davidson A, Wood AJ, Li G, Sun LS. Long-term differences in language and cognitive function after childhood exposure to anesthesia. Pediatrics 2012;130:e476-485.

**Table 1.** Patient Demographics

| Characteristics | No. Patients |
|---|---|
| Total | 37 |
| Mean age at presentation (days ± SD) | 26 ± 60 |
| Mean age at electrodiagnostic testing (days ± SD) | 41 ± 60 |
| Median Narakas score | 3 |
| Narakas I to II* | 21 (57%) |
| Narakas III to IV* | 16 (43%) |
| Gender | |
| Boy | 21 (57%) |
| Girl | 16 (43%) |
| Race | |
| Caucasian | 22 (59%) |
| African American | 3 (8%) |
| American Indian/Alaska Native | 1 (3%) |
| Other/unknown | 11 (30%) |
| Lesion side | |
| Left | 20 (54%) |
| Right | 16 (43%) |
| Both | 1 (3%) |

*Narakas classification is a reflection of the extent of nerve root involved. Narakas I represents involvement of C5 and C6 nerve roots; Narakas II, C5-C7; Narakas III, C5-T1; Narakas IV, C5-T1 with Horner's syndrome.

**Table 2.** Motor and Sensory Nerve Conduction Studies Performed in the Study Cohort (N = 37)

| | Unilateral<br>n (%) | Bilateral<br>n (%) | None<br>n (%) |
|---|---|---|---|
| Median-sensory | 32 (86) | 1 (3) | 4 (11) |
| Ulnar-sensory | 30 (81) | 1 (3) | 6 (16) |
| Median-motor | 23 (62) | 1 (3) | 13 (35) |
| Ulnar-motor | 26 (70) | 0 | 11 (30) |

**Table 3.** Frequencies of Muscles Examined by Needle Electromyography in the Study Cohort (N = 37)

| Nerve Root/Muscle | Muscles N (%) |
|---|---|
| C5 | |
| Biceps | 34 (92) |
| Deltoid | 34 (92) |
| Infraspinatus | 1 (3) |
| C6 | |
| Biceps | 34 (92) |
| Brachioradialis | 1 (3) |
| Extensor carpi radialis | 1 (3) |
| C7 | |
| Triceps | 32 (86) |
| Flexor carpi radialis | 3 (8) |
| Extensor digitorum communis | 0 (0) |
| C8 | |
| Flexor digitorum superficialis | 1 (3) |
| Extensor indicis | 2 (5) |
| Extensor digitorum communis | 6 (16) |
| T1 | |
| First dorsal interosseous of hand | 15 (41) |
| Abductor pollicis brevis | 1 (3) |

**Table 4.** Percentage Agreement and Interrater Reliability Between Examiners, by Nerve Roots

| Nerve Root | Agreement n (%) | Disagreement n (%) | Kappa | 95% CI | *P*-Value* | Status |
|---|---|---|---|---|---|---|
| C5 | 14 (38) | 23 (62) | 0.27 | 0.16 - 0.38 | < 0.0001 | Fair |
| C6 | 29 (78) | 8 (22) | 0.65 | 0.44 - 0.86 | < 0.0001 | Substantial |
| C7 | 34 (92) | 3 (8) | 0.86 | 0.71 - 0.99 | < 0.0001 | Almost perfect |
| C8 | 30 (81) | 7 (19) | 0.65 | 0.42 - 0.88 | < 0.0001 | Substantial |
| T1 | 31 (84) | 6 (16) | 0.73 | 0.54 - 0.93 | < 0.0001 | Substantial |
| All | 138 (75) | 47 (25) | 0.62 | 0.53 - 0.72 | < 0.0001 | Substantial |

*Bonferroni correction for 37 subjects to maintain an evaluation-wise error rate of 0.05; only *P*-values <0.001 were considered statistically significant in this analysis.

**Table 5.** Interrater Reliability Between Examiners, by Nerve Trunks

| Nerve Trunk | Agreement n (%) | Disagreement n (%) | Kappa | 95% CI | *P*-Value* | Status |
|---|---|---|---|---|---|---|
| Upper | 43 (58) | 31 (42) | 0.43 | 0.27 - 0.58 | < 0.0001 | Moderate |
| Middle | 34 (92) | 3 (8) | 0.86 | 0.71 - 0.99 | < 0.0001 | Almost perfect |
| Lower | 61 (82) | 13 (18) | 0.70 | 0.55 - 0.85 | < 0.0001 | Substantial |
| All | 138 (75) | 47 (25) | 0.62 | 0.53 - 0.72 | < 0.0001 | Substantial |