RELIABILITY OF INFANIB AMONG NORMAL AND HIGH RISK TERM NEONATES: A CROSS SECTIONAL STUDY

MY Wong[1], Sai L Ganesan[2], UV Shenoy[3], *Jaya S Tedla[4], VS Binu[5]

ABSTRACT

Background: *Infant Neurological International Battery (INFANIB) is one of the common neurological assessment used for the infants. The reliability in the newborns is more essential to prove because INFANIB is used to predict the infant motor capabilities at later stages. Objective: Objective of the study is to establish intra-rater and inter-rater reliability of Infant Neurological International Battery between normal and high risk term neonates.* ***Methods***: *In this cross sectional study a convenience sample of term neonates was included. Demographic data were completed from medical records. Rater one administered INFANIB on all the neonates and the assessments were being videotaped. The videos were anonymized and were viewed by two independent raters so they were blinded from knowing the groups from which each neonate belonged to. Rater three viewed 50 videos randomly for assessment of intra-rater agreement after one month. All the raters were blinded from scores of one another as well to avoid bias during assessment. For intra- as well as inter-rater reliability, two-way mixed, absolute agreement and single measure ICC was calculated.* ***Results***: *112 neonates (normal=56; high risk=56) with mean (SD) gestational age of 39.1 (1.1) and 38.9 (1.1) weeks respectively on mean post natal day of 4 were assessed. For intra-rater agreement the ICC was 0.42 (0.05-0.7) for normal neonates and 0.33 (0-0.63) for high risk neonates. For inter-rater agreement the ICC was 0.1 (0-0.26) for normal neonates and 0.28 (0.05-0.5) for high risk neonates.* ***Conclusions***: *The intra-rater and inter-rater reliability of INFANIB proved to be poor between the normal and high risk term neonates. Formal training should be given to any profession or health workers who are planning to use INFANIB as a screening tool for early neurodevelopmental delay among neonates to improve the inter-rater agreement.*

KEY WORDS: *Neonatology, Neurodevelopment, Reliability, INFANIB, Screening.*

ملخص: البطارية العصبية الدولية للأطفال الرضع (INFANIB) هي واحدة من اكثر النظم شيوعا لتقييم الجهاز العصبي للأطفال الرضع, و من الاهمية اثبات موثوقية هذه الوسيلة في الكشف علي الأطفال حديثي الولادة لأنه يتم استخدام هذا النظام للتنبؤ بالمقدرات الحركية للرضع في مراحل لاحقة. **الهدف:** الهدف من هذه الدراسة هو تحديد موثوقية نظام التقييم بين مختلف المقيمين الذين يقيمون حديثي ولادة طبيعيين واخرين ذو معدل خطورة عالي. **منهج البحث:** هي دراسة مقطعية اشتملت علي عينة مناسبة من الاطفال حديثي الولادة كاملي النمو. وقد اخذت البيانات الديموغرافية من السجلات الطبية ثم تم تقييم كل الاطفال بواسطة المقيم الاول و جري تصوير عملية الكشف والتقييم بالفيديو. روجعت الفيديوهات من قبل مقيمين اخرين دون معرفة لأي فئة يتبع الاطفال. بعد شهر من التقييم تم عرض ٥٠ فيديو عشوائيا علي مقيم ثالث لمعرفة التوافق في النتائج بين المقيم الاول والثاني. كل مقيم ادي دوره دون معرفة نتائج الاخر لتفادي الانحياز . لحساب الموثوقية بين المقيمين تم التحليل باستخدام معامل الارتباط (Two-way mixed, absolute agreement and single measure ICC).

Received: 21 March, 2013; Accepted: 2 June, 2013

**Correspondence*: shankar_tedla@yahoo.co.in
[1]Department of Paediatrics, Kasturba Medical College, Mangalore, India; [2]Ex-Professor, Department of Physiotherapy, Kasturba Medical College, Mangalore, India; [3]Professor, Department of Paediatrics, Kasturba Medical College, Mangalore, India; [4]Assistant Professor, Program of Physical Therapy, Department of Medical Rehabilitation Sciences, College of Applied Medical Sciences, King Khalid University, Abha, Kingdom of Saudi Arabia; [5]Senior Grade Lecturer, Department of Statistics, Manipal University, Manipal, India.

النتائج: تم التقييم لعدد ١١٢ طفل حديث ولادة (٥٦ طبيعي و ٥٦ ذو معدل خطورة عالية) وقد كان التقييم (في المتوسط) في اليوم الرابع للولادة. متوسط العمر الحملي ٣٩،١ أسبوع مع انحراف معياري ١،١ للطبيعيين و ٣٨،٩ أسبوع مع انحراف معياري ١،١ لذوي الخطورة العالية . معامل الارتباط بين المقيمين الاول والثاني كانت ٠،٤٢ (٠،٠٥ حتي ٠،٧) لحديثي الولادة الطبيعية و ٠،٣٣ (٠-٠،٦٣) لحديثي الولادة ذو المخاطر العالية. معامل الارتباط بين المقيم الثالث والاخرين كانت ٠،١ (٠-٠،٢٦) لحديثي الولادة الطبيعية و ٠،٢٨ (٠،٠٥-٠،٠٥) لحديثي الولادة ذو المخاطر العالية. **الاستنتاجات**: اثبتت الدراسة ضعف الموثوقية في تطبيق نظام التقييم بواسطة المقيمين المختلفين بين حديثي الولادة الطبيعيين وذوي المخاطر العالية , مما يستوجب التدريب الرسمي للعاملين في المجال الصحي الذين يستخدمون البطارية الدولية للأطفال كأداة فحص أولي لتشخيص تأخر النمو العصبي في وقت مبكر بين حديثي الولادة.

INTRODUCTION

Neonatal period is a crucial period in detecting neurological illness.[1] Various studies have shown the value of early identification of developmental delay among children.[2,3] Following which timely intervention can be given, these include physical therapy, speech and language therapy, occupational therapy and sensory integration. Evidence showed that early treatment of developmental delay leads to better outcomes for children and reduces the costs of providing health care, educational support and treatment services. Hence, there is an increasing emphasis on the use of appropriate developmental screening for children and early detection of neurodevelomental delay. The ideal screening method should be a standardized and validated tool with established psychometric properties, easy to perform and interpret, cost effective, and have a good sensitivity and specificity.[2]

Infant Neurological International Battery (INFANIB) was designed to assess the motor development of infants whose corrected gestational age ranged from 0 to 18 months though INFANIB was constructed based on a study consisted of infants from age three to 22 months in 198.[4,5] In view of motor development that is most rapid in infancy,[6] INFANIB is a screening tool used for early detection of gross motor developmental delay.[7] INFANIB has 20 items which were selected from four methods of neurological examination: 1) the Milani-Comparetti and Gidoni method, 2) the French "angles" method, 3) the primitive reflexes method, and 4) the Paine and Oppe method[8].

INFANIB is an instrument to assess the neurological integrity of infants and used in follow-up programs. The assessment is practical because it can be completed by the clinicians within several minutes.[4,8] In view of tone abnormalities detected in early infancy may resolve at the end of the first year of life,[9,10] the standard scoring sheet of INFANIB, which allows clinicians and physical therapists to evaluate the infant's tone and posture in a quantified manner[11] give a more accurate diagnosis when the child is being followed-up in the future.

INFANIB was constructed based on practicality and psychometric principles, but reliability is the major concern in view of the variation in the force that the examiner applies to the limbs of the infant and variation in the measurement of angle. Occasionally, clinicians and physical therapists have different rating for the same infant in view of different interpretations of the results when similar methods of examination have been used. This is due to the fact that clinicians examined the infants to make a diagnosis while physical therapist identified minor impairments for therapy.[4] There is scarcity of literature available on the inter- and intra-rater reliability of the INFANIB scale in the neonates.[3,4,8] Since early detection of high risk infant is crucial

for further follow up in the future, neonatal group was chosen as the subject of interest.

MATERIALS AND METHODS

This cross sectional study included a total of 112 term neonates (56 Normal and 56 High Risk) in the state 3 (eyes open, no gross movement) and 4 (eyes open, gross movements, no crying) of Normal Neonatal Behavioral States of Prechtl and Beintema.[12] This study was conducted in the year 2010-2012 and the full term newborns were selected by convenient sampling as and when they are available in the hospital. Preterm neonates were excluded to eliminate errors from hypotonia of prematurity.[11] Term neonates with systemic illness and congenital musculoskeletal disorder were not eligible. In addition, term neonates under maternal medication in whom alertness was affected were excluded.[13] High risk term neonates were those neonates with risk factors associated with neurodevelopmental delay. The risk factors included low birth weight, perinatal asphyxia, low Apgar score, neonatal hypoglycemia, neonatal seizures, type II pneumonia, and pregnancy complications.[3,12,14,15]

Approval from Institutional Ethical Committee and permission from the Medical Superintendent of the Government Hospital were obtained prior to commencement of the study. Informed signed consents were taken from the parents. The demographic data were completed using the medical records. The INFANIB was administered on the neonates by Rater one, an Intern of MBBS who was new to INFANIB and studied the method of assessment by reading the INFANIB manual written by Patricia H. Ellison, M.D. and via several demonstrations by the experienced physiotherapist. During each assessment, a video recording was taken simultaneously.

The videos were being anonymized to blind the raters from the groups from which the neonates belonged. The videos were viewed by two other independent raters they are rater two and rater three for the purpose of assessing inter rater reliability of the scale.

Rater two was a postgraduate from the Physiotherapy Department who was inexperienced in the use of INFANIB scale and Rater three was an experienced physiotherapist in pediatrics and was skilled in neurological evaluation of infants using INFANIB scale. Rater three had evaluated the score twice for the purpose of intra-rater reliability. All the raters were blinded to the scoring of one another to avoid bias.

Instrument

14 out of 20 items in the INFANIB scale were applied to the neonates. Each item was scored as 1, 3 or 5 points. The cutoff points for neonatal category was as follows: abnormal $\leq$ 48, transient = 49-65, normal$\geq$ 66.[8,11]

Statistical analysis: Statistical analyses were performed using SPSS version 11.5.0. Gestational age of the neonates, postnatal day on which they assessed and birth weight was analyzed by using descriptive statistics. Two-way mixed, absolute agreement and single measure ICC were calculated for intra-rater reliability as well as inter-rater reliability.

RESULTS

A total of 112 term neonates consisting of 56 normal (25 males and 31females) neonates and 56 high risk (31 males and 25 females) neonates was assessed by three raters in this study. Rater three employed INFANIB scale to all the neonates and 50 videos were randomly selected (23 normal and 27 high risk neonates) for reassessment after one month to assess intra-rater reliability.

Perinatal characteristics

The mean (SD) gestational ages of neonates in normal and high risk group were 39.1 (1.1) weeks and 38.9 (1.1) weeks respectively. The median and quartiles for a post natal day on which they were being evaluated were 4 (2 to 5) days and 4 (2.3 to 7) days respectively. The mean (SD) birth weight of normal neonates was 2.9 (0.3) kilograms and for high risk group was 2.7 (0.6) kilograms. Risk factors included in this study were prenatal history of two or more abortion, pre-eclampsia, maternal illness, gestational diabetes mellitus, oligo/polyhydramnios, natal history of meconium stained amniotic fluid, prolonged rupture of membranes, prolonged labor, cephalopelvic disproportion and post natal history of resuscitation, low Apgar scores (<3 at 1 minute),[12] low birth weight (birth weight <2500g),[7] neonatal sepsis and pathological neonatal hyperbilirubinemia.

Reliability	Group	ICC	95% CI
Intra-rater (Rater three)	Normal (n=23)	0.42	0.05-0.7
	High risk (n=27)	0.33	0-0.63
Inter-rater (Rater one-Rater two - Rater three)	Normal (n=56)	0.1	0-0.26
	High risk (n=56)	0.28	0.05-0.5

Reliability of INFANIB

Intra-rater agreement the ICC was 0.42 (0.05-0.7) for normal neonates and 0.33 (0-0.63) for high risk neonates. Inter-rater agreement the ICC was 0.1 (0-0.26) for normal neonates and 0.28 (0.05-0.5) for high risk neonates. The intraclass correlation coefficient for the intra-rater agreement of the Rater 3 and inter-rater agreement are shown in Table-1.

DISCUSSION

During the construction of INFANIB, Ellison et al had done three hundred and sixty-five assessments among infants between 3 and 22 months old. The study concluded that the internal consistency of this scale was 0.91 for all subjects. For infants aged less than 7 months, the internal consistency of INFANIB scale was 0.88 and for infants aged 8 months or more it was 0.93.[8]

According to Soleimani et al, the reliability study between the pediatrician and occupational therapists was conducted among 6150 infants aged from 4 to 18 months. The intraclass correlation coefficient was 0.90.[7]

In the current study, the intra- and inter-rater reliability of INFANIB was poor among the raters in both groups of neonates. The raters in this study have different level of experience in the use of INFANIB. The purpose of this study was to find out if there was good inter-rater agreement among different professionals and health workers so that it could be used more widely. Hence, a pilot study for interpretation of the finding for each item in INFANIB was not done for Rater one by skilled professionals in order to find out whether it could be used widely by the health workers in the peripheral area as a screening tool without formal training beforehand.

Only term neonates in the state 3 and 4 of Normal Neonatal Behavioural States of Prechtl and Beintema[12] were included in this study due to the fact that the tone of neonates changes according to the neonatal state which would misguide the raters to evaluate hypertonia if they were in state five and hypotonia if they were in state one and two. Furthermore, term neonates usually sleep 50 minutes out of each hour of the day[12] and this prevented each rater to assess the neonates individually. Hence the use of video capture was chosen. Intra- and inter-rater disagreement between the raters was not due to the variation of tone in the neonates since they rated the individual

neonate from the same video. In addition, use of video capture for the study was decided from the ethical point of view since repeated handling by different raters would carry the risk of infection to the susceptible neonates and there was a need for neonates being by their mothers' side for warmth and breastfeeding. Therefore, the poor intra- and inter-rater agreement was purely affected by the interpretation of findings by the raters.

The lack of pictorial representation in item "hands closed/open" in the INFANIB scoring sheet made it difficult for the inexperienced professional in making the decision of the findings. Most of the neonates had their hands partially closed and the only options available were "Closed" and "Sometimes closed". The experienced professional agreed on "Sometimes closed" (score 3) while the new user of INFANIB interpreted it as "Closed" (score 5).

Interpretation of items in the French angle method[11] such as scarf sign, heel to ear, popliteal angle, leg abduction and dorsiflexion of foot was very subjective. The raters analyzed their observation according to their imaginary protractors. Although it may not be feasible, the use of large protractor or goniometer in the clinical setting may improve the intra- and inter-rater reliability since it is more objective.[16] The evaluation of the above items also required the examiner to feel the resistance during the movement which was crucial in judging the range of angles. However, it was not possible for each of the raters to assess each neonate in view of ethical issue and the risk of infection to the neonates.

In items "tonic labyrinthine-supine and prone"11, there was a need to observe the shoulders, trunk and legs simultaneously which would have been difficult for the less experienced raters. They could have missed the findings which led to poor inter-rater agreement.

For the item "weight bearing", the new user could not detect the difference between "No weight bearing"11 (score 5) and "Poor weight bearing, breaks at knees"11 (score 3). This was because the neonates could hardly stand for a few seconds before bending their knees in the latter. It was those few seconds which differentiated whether the neonates would be scored as "3" or "5".

Manual review of data for individual items of all the raters revealed maximum agreement in item "pull to sitting"11 and good agreement in items "sitting position"11 and "asymmetric tonic neck reflex"(ATNR). This could be due to the clear pictorial representation of findings in items "pull to sitting" and "sitting position" in the INFANIB scoring sheet that simplified the interpretation of findings among the raters. Meanwhile, item "ATNR" was a primitive reflex with clear cut findings which could be easily identified by the raters. It could only be "persistent", "transitional" or "absence" of posturing.

In the item "lifts head"11, there were several neonates who did not raise their heads but there were no results for "no head lifting" in the scoring sheet. In this study, we synchronize it as "3 scores" for this finding.

The poor inter-rater reliability found in this study could also be explained by the fact that INFANIB was designed for the professionals including the pediatricians and physical therapists. On top of that, lack of formal training beforehand made the less experienced raters unable to interpret the findings confidently. Even though the recent study on the Chinese version of INFANIB have excellent reliability for 3, 7 and 10 months old infants but reliability in

neonates is still needs to be explored further.[17]

The use of video limited the raters from judging the force exerted by the neonates in item "foot grasp" as well as items of "French angles method". This contributed to the poor intra-rater reliability in the interpretation of findings by the experienced rater. Due to convenient sampling, results may not be generalized and the present study was limited in the number of neonates aged more than 7 days. There were only 7 and 9 neonates in normal and high risk groups respectively who were aged from 8 to 21 days post natally. Future studies should fill in the gap of day 8 to 28 in neonatal groups. In addition, the inter-rater reliability between the experienced raters can also be carried out in the future.

In conclusion, the intra- and inter-rater reliability are poor among the normal and high risk term neonates. However, INFANIB is potentially suitable and easy tool to screen for early neurodevelopmental delay among the term neonates because it can be completed in 5 to 10 minutes. Therefore, formal training should be organized for any professions or health workers who are planning to utilize INFANIB in their working set up to improve the inter-rater agreement.

ACKNOWLEDGEMENT

All the authors thank Miss Anupama Bajaj, postgraduate in physiotherapy for her participation in video rating. We are grateful to Dr Shetty Subhodh, Dr Sarfaraz Nawas, nursing faculty in the Government Hospital and colleagues in their participation during the data collection. We also thank Dr. B. Unnikrishnan and Mr. Shashidhar Kotian for their cooperation and guidance.

REFERENCES

1. Rennie JM. Roberton's Textbook of Neonatology. 4th Ed. Edinburgh(UK): Elsevier Churchill Livingstone; 2005. p. 1093-105.
2. Poon JK, Rosa ACL, Pai GS. Developmental Delay: Timely Identification and Assessment. Indian Paed 2010;47:415-22.
3. Soleimani F, Vameghi RV, Hemmati S, Roghani RS. Perinatal and Neonatal Risk Factors for Neurodevelopmental Outcome in Infants in Karaj. Arch Iran MED 2009;12 (2): 135-9.
4. Ellison PH. The INFANIB, A Reliable Method for the Neuromotor Assessment of Infants . Therapy Skill Builder, Tucson, AZ, USA. 1994
5. Wilhelm IJ. Physical therapy assessment in early infancy. Churchill Livingstone, New York, USA. 1993: 46-8.
6. Lissauer T, Clayden G. Illustrated Textbook of Paediatrics. 3rd Ed. Edinburgh (UK): Mosby Elsevier; 2007. p. 23-38.
7. Soleimani F, Dadkhah A. Validity and reliability of Infant Neurological International Battery for detection of gross motor developmental delay in Iran. Child Care Health Dev 2007;33:262-5.
8. Ellison PW, Horn JL, Browning CA. Construction of an Infant Neurological International Battery (INFANIB) for the assessment of neurological integrity in infancy. Phys Therapy 1985;65:1326-31.
9. Amiel-Tison C, Grenier A. Neurological assessment during the first year of life. Oxford University Press, New York, USA. 1986.
10. Chaudhari S, Bhalerao M, Chitale A, et al. Transient tone abnormalities in "high risk" infants and cognitive outcome at five years. Indian Pediatr 2010;47:931-5.
11. Ellison PW. Scoring Sheet for the Infant Neurological International Battery (INFANIB). Phys Ther 1986;66:548-50.
12. McIntosh N, Helms PJ, Smyth RL, Logan S. Forfar & Arneil's Textbook of Pediatrics. 7th Ed. Edinburgh: Churchill Livingstone Elsevier; 2008. p. 297-301.
13. Van Baar AL, Fleury P, Soepatmi S, et al. Neonatal behavior after drug dependent pregnancy. Arch Dis Child 1989;64:235-40.
14. Reddihough DS and Collins KJ. The epidemiology and causes of cerebral palsy. Aust J Physiother 2003;49:7-12.

15. Elenjickal MG, Thomas K, Sushamabai S, Ahamed SKZ. Development of high-risk newborns-A follow-up study from birth to one year. Indian Pediatr. 2009;46:342-5.
16. Chaudhari S, Deo B. Neurodevelopmental assessment in the first year with emphasis on evolution of tone. Indian Pediatr 2006;43:527-34.
17. Liao W, Wen EY, Li C, Chang Q, Lv KL and Yang W etal. Predicting neurodevelopmental outcomes for at-risk infants: reliability and predictive validity using a Chinese version of the INFANIB at 3, 7 and 10 months. BMC Pediatr. 2012 ;12:72. doi: 10.1186/1471-2431-12-7.