Published in final edited form as:

J Pediatr. 2007 June ; 150(6): 592–596.e5. doi:10.1016/j.jpeds.2007.02.012.

INTEROBSERVER RELIABILITY AND ACCURACY OF CRANIAL ULTRASOUND INTERPRETATION IN PREMATURE INFANTS

Susan R. Hintz, Thomas Slovis, Dorothy Bulas, Krisa P. Van Meurs, Rebecca Perritt, David K. Stevenson, W. Kenneth Poole, Abhik Das, and Rosemary D. Higgins [on behalf of for the NICHD Neonatal Research Network]^{*}

Abstract

Objective—To assess interobserver reliability between two central readers of cranial ultrasound (CUS) and accuracy of local compared with central interpretations.

Study design—A retrospective analysis of CUS data from the NICHD trial of inhaled nitric oxide for premature infants. Interobserver reliability of two central readers was assessed by kappa or weighted kappa. Accuracy of local compared with central interpretations was assessed by sensitivity and specificity.

Results—Cranial US from 326 infants had both central reader and local interpretations. Central reader agreement for grade 3/4 IVH, grade 3/4 IVH or PVL, grade of IVH, and degree of ventriculomegaly was very good (kappa=0.84, 0.81, 0.79, and 0.75, respectively). Agreement was poor for lower grade IVH and for PVL alone. Local interpretations were highly accurate for grade 3/4 IVH or PVL (sensitivity 87–90%, specificity 92–93%), but sensitivity was poor to fair for grade 1/2 IVH (48–68%) and PVL (20–44%).

Conclusions—Our findings demonstrate reliability and accuracy of highly unfavorable CUS findings, but suggest caution when interpreting mild to moderate IVH or white matter injury.

Keywords

intraventricular; hemorrhage; leukomalacia; central reader; neurodevelopmental; brain

Cranial ultrasound (CUS) is one of the most important diagnostic procedures performed in the neonatal intensive care unit (NICU). After Pape, et. al. (1) reported bedside ultrasound to detect intraventricular hemorrhage (IVH) in premature infants, it quickly became the neuroimaging standard of care (2–4). Periventricular leukomalacia (PVL) can be detected by CUS as cystic lesions or, in evolution, as periventricular echodensities (5). Outcome studies of preterm infants

© 2007 Mosby, Inc. All rights reserved.

Address correspondence to: Susan R. Hintz, M.D., M.S. (Epi), Assistant Professor of Pediatrics, Division of Neonatal and Developmental Medicine, Stanford University School of Medicine, 750 Welch Road, Suite 315, Palo Alto, CA 94304, Phone: 650-723-5711, Fax: 650-725-8351, email: srhintz@stanford.edu.

Edited by WFB

^{*}List of members of the NICHD Neonatal Research Network is available in Appendix I at www.jpeds.com

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.

Disclosure: INO Therapeutics provided the study gas and gas delivery systems for all hospitals, and capitation funding for the hospitals outside the NICHD Neonatal Research Network that participated in the PiNO trial. The company was not involved in the trial design, data analysis or interpretation, or secondary or ancillary data analysis. The authors have no financial agreement with INO Therapeutics. For funding information, see Appendix II at www.jpeds.com.

have revealed strong associations of severe CUS abnormalities with later major neurodevelopmental disabilities (6–10).

Despite the apparently important role of accurate CUS interpretation, few studies have investigated reliability or accuracy. Randomized controlled trials and interventional studies involving premature infants often include severe CUS abnormalities among study outcomes, only occasionally using central reader interpretations. Observational studies routinely report frequencies of abnormal CUS findings based solely on local interpretations. However, although intraobserver reliability analyses of CUS interpretations have been previously reported (11–14), little attention has been given to central reader reliability or to accuracy of local compared with central reader interpretations.

Substantial central and local reader CUS data are available from the multicenter NICHD Neonatal Research Network randomized, controlled, double-masked trial of inhaled nitric oxide (iNO) for severe respiratory failure in premature infants (PiNO trial) (15). We undertook an analysis of these data to assess interobserver reliability between two central readers, and to assess accuracy of local interpretations compared with central readers.

METHODS

Study design and population

This was a retrospective reliability and accuracy study of CUS data from the PiNO trial (15) (infants less than 34 weeks gestation and 401–1500 g), and from the concurrently enrolled larger premature infant pilot (infants less than 34 weeks but greater than 1500 g). A secondary hypothesis of the trial was that surviving infants in the iNO group would have no increase in grade 3 or 4 IVH or PVL compared with those in the placebo group. A single CUS was originally required at 28 ± 3 days of age among survivors, which was to be read by two central readers. However, after interim analysis, the Data Safety and Monitoring Committee recommended that copies of all CUS performed during hospitalization be requested from sites for central reading. Data collection regarding local interpretations was nevertheless required only for the 28 ± 3 day CUS, if the patient survived to that time. Data collection of local interpretations of other CUS was requested, but not required, for three other specifically timed CUS, if they were performed: before study gas administration, during study gas administration, and after 28 ± 3 days of age.

Thus, for the current analysis, a single CUS was included from each patient for whom a CUS was submitted and for which both local and central readings were documented. We structured our analysis in this way in part to limit potential bias due to central reader misclassifications carried through an entire patient CUS series. The 28-day CUS was used if it existed; however, because approximately 50% of these critically ill infants died, the CUS closest to that time point was used if a 28-day study was not obtained.

Local reader CUS interpretations: PiNO trial data collection and definitions

Trained research staff at each institution collected data regarding local CUS interpretations directly from local radiologists' reports. The PiNO trial did not limit the number of local radiologists or technologists involved. Cranial US were obtained and read per local institution clinical protocol; the local site radiology departments were not required to adhere to policy with respect to CUS view acquisition, nor did local radiologists implement any changes to their interpretation approach due to the PiNO trial. The PiNO trial manual of operations provided suggested definitions for grades of IVH (Grade 1: Hemorrhage confined to the germinal matrix/subependymal area; IVH Grade 2: Hemorrhage in the lateral ventricle(s) without distension; Grade 3: Hemorrhage in the lateral ventricle(s) with distension; Grade 4: Hemorrhage

extending to the brain parenchyma), but no special training was undertaken for local site radiologists or technologists with respect to the PiNO trial. Echodense PVL was not distinguished from echolucent PVL on local interpretation data queries; PVL was coded as present if the local radiologist final clinical report documented PVL on CUS. Research staff at each institution completed standardized data collection forms from the local radiologist final reports, and transferred data by secure computer network to the data center (RTI International, Research Triangle Park, N.C.). No queries with respect to side of lesion were included; thus, IVH on either or both sides was coded as “Yes”. If IVH was present bilaterally, the most severe IVH was coded. Data with respect to ventriculomegaly were not collected on the local interpretation data collection instrument.

Central reading procedures and definitions

Copies of CUS were sent by centers to RTI International in film or digital format. Two central readers reviewed CUS during a two-day period after the trial closed. The central readers were board-certified pediatric radiologists from different academic institutions, and had special expertise in CUS. Prior to the trial, the central readers collaborated with the PiNO trial subcommittee to create the central reader data instrument (Figure; **available at** www.jpeds.com), which collected detailed, hemisphere-specific radiologic observations and diagnostic classifications. They briefly reviewed the instrument together before the central reading session, but interpreted CUS independently. For the PiNO trial (15), a third reader later adjudicated discrepant interpretations between the two central readers, but this was limited only to those studies in which the global CUS diagnosis of grade 3 or 4 IVH or PVL on either hemisphere was not agreed upon.

Analysis Plan

Given differences between the central and local reader data collection instruments, reliability and accuracy analyses focused on important clinical diagnostic and prognostic category queries common to both. Because local data collection queries did not differentiate PVL type, we limited both reliability and accuracy analyses to any PVL. Only central reader reliability analyses could be performed with respect to ventriculomegaly because local data were not collected. SAS 9.1.2. (SAS Institute, Inc., Cary, NC, USA) was used for all analyses.

Reliability analyses: Agreement between central readers 1 and 2—Assessment of reliability between central readers was by kappa statistic. Kappa measures the extent of agreement beyond that which would be expected by chance alone under the assumption of independence, described as:

$$\text{Kappa} = \frac{\% \text{ observed agreement} - \% \text{ expected agreement}}{100\% - \% \text{ expected agreement}}$$

Several kappa interpretation schemes have been reported (16–19), with kappa >75% generally considered “good” to “excellent”.

We tested central reader agreement by kappa on dichotomous CUS observations, including normal reading, grades of IVH, any PVL, grade 3 or 4 IVH or PVL, and any ventricular size increase. Multicategory weighted kappa analysis (18,20,21) was performed for grade of IVH and severity of ventriculomegaly. The weighted kappa allows for some credit to be given for minor misclassifications (20,21). Because kappa is limited by its dependence on prevalence (22,23), use of accompanying measurements of agreement such as percent positive agreement (PPA) have been recommended (24).

Accuracy analyses: Comparing local CUS interpretations to “gold standard” central readers—Assessment of accuracy was by sensitivity and specificity analyses, using central reader interpretation as the “gold standard” reference. Analyses separately compared local interpretations to central reader 1 (CR #1) and central reader 2 (CR #2). Sensitivity was defined as the proportion “positive” by local among all “positive” by central reader; specificity was defined as the proportion “negative” by local among all “negative” by central reader. Sensitivity and specificity analyses evaluated diagnostic categories common to both the local reader queries and central reader forms. Because instruments differed with respect to whether CUS findings were recorded as global or hemisphere-specific, the grade of IVH recorded on the local reader form was considered “correct” if the highest grade IVH recorded on either hemisphere by the central reader was the same as the local reader result.

Ethical Considerations

The Institutional Review Boards (IRB) of all centers reviewed and approved of the study. Informed consent was obtained from the patient’s parent or legal guardian prior to randomization to iNO or placebo. Patient identifiers, including name and hospital medical record number, were removed by hand or digitally from CUS studies prior to being sent to RTI for central reading.

INO Therapeutics (Clinton, New Jersey) provided the study gas and gas delivery systems for all hospitals, and capitation funding for the hospitals outside the NICHD Neonatal Research Network that participated in the PiNO trial. The company was not involved in the trial design, data analysis or interpretation, or secondary or ancillary data analysis.

RESULTS

449 infants were enrolled in the PiNO trial and larger premature infant pilot; 70 had no CUS performed. Of those remaining, 361 patients had at least one CUS of adequate quality to result in completed central readings, and 326 CUS had both central and local readings. Observations for all diagnostic categories were not available due to missing or uninterpretable data entries.

Reliability analyses

326 CUS studies and 650 hemispheres were read by both central readers. The diagnostic assignments made by the central readers are shown in Table I.

Kappa and weighted kappa analyses—Central reader reliability for major diagnostic categories is shown in Table II. The best central reader agreement was achieved for the major adverse prognostic categories, IVH grade 3 or 4 and IVH grade 3 or 4 or PVL. Agreement between central readers was poor for lower grade IVH, and for PVL alone.

When complete IVH grade classification data (Table III; **available at** www.jpeds.com) is viewed with respect to prognostic categorization, the results reflect very good agreement (multilevel weighted kappa=0.79, 95% CI 0.75–0.82). Of 426 hemispheres coded by CR #1 as no IVH, 3 were coded by CR #2 as grade 4 IVH; of 391 hemispheres coded by CR #2 as no IVH, 3 were coded by CR #1 as grade 4 IVH. In 2 of the 6 cases, central readers agreed that intracerebral, but not intraventricular or periventricular echodensities, were present. In 3 cases, the central readers agreed that grade 3 or 4 IVH or PVL existed on the contralateral hemisphere, thus the third reader did not adjudicate those discrepancies. In one case, the adjudicator considered the study to be of poor quality and uninterpretable.

The complete ventricular enlargement classification data (Table IV; **available at** www.jpeds.com) are also reassuring. Only 2 of the 516 ventricles coded by CR#1 as no or mild

enlargement were coded as severely enlarged by CR#2. None of the 488 ventricles coded by CR#2 as no or mild enlargement was coded as severely enlarged by CR#1.

Accuracy analyses: Local readers compared with central readers

Sensitivity and specificity of local reader interpretations—Table V summarizes sensitivity and specificity for local interpretations compared with CR #1 and CR #2 for major diagnostic and prognostic categories. These results demonstrate excellent sensitivity and specificity for local interpretations for broad diagnostic categories, and for the major adverse prognostic category of grade 3 or 4 IVH or PVL. However, sensitivity was poor for lower grades of IVH and for PVL. Complete classification tables for IVH grades and major diagnostic category interpretations for local compared with central reader interpretations are shown online (Table VI–Table VIII; **available at** www.jpeds.com).

DISCUSSION

In this analysis of CUS interpretations of PiNO trial subjects, we found that reliability between central readers was excellent for severe IVH, the major prognostic category of grade 4 or 4 IVH or PVL, and degree of ventriculomegaly, but was substantially poorer for grade 1 or 2 IVH. Local CUS interpretations were highly sensitive and specific for severe hemorrhage and for global presence of IVH, but sensitivity was poorer for lower grades of IVH. Our results also reflect significant variability in interpretation of PVL by CUS. These findings are reassuring with respect to highly unfavorable hemorrhagic CUS findings associated with adverse neurodevelopmental outcome, but suggest caution when interpreting reports of mild to moderate hemorrhage. Advanced neuroimaging methods such as MRI would better determine the extent of subtle brain injury and white matter injury.

Our interobserver analyses compare favorably with the few previous detailed studies. Pinto et al. (12) also found that agreement for severe findings was excellent, but much worse for less severe categories. Although the Pinto study reported generally better agreement than ours, the radiologists and technologists had training sessions prior to and during the study. The study by Corbett et al (13) also showed very good interobserver agreement for severe findings, but poor for lower IVH grades and ventricular enlargement. No special training or certification process was undertaken as part of the Corbett study, thus, our analysis and the Corbett study may better reflect the “true” level of agreement between independent radiologists at different institutions.

Our reliability analyses reveal poor agreement on PVL between central readers. The prevalence of PVL was low, thus agreement analyses by either kappa or PPA may not be particularly informative. The clinical importance of the poor agreement in this scenario is also unclear. Echodense PVL is often transient, and the prognostic significance of a single CUS with echodense PVL may be minimal (25). Nevertheless, this disagreement on CUS interpretations of white matter injury, as well as the low rate of PVL in this study compared with that recognized by MRI, reinforces the potential substantial diagnostic and predictive benefit afforded by modalities such as MRI (26).

Very few hemispheres interpreted by one central reader as no IVH were interpreted as grade 4 by the other. In half of these cases, the global CUS categorization to an adverse prognostic category would not have changed because the contralateral hemisphere was affected by severe IVH or PVL. In some instances, pure parenchymal hemorrhages appear to have been variably coded as no IVH and grade 4 by the two central readers. The quality of CUS film or digital media copies from multiple sites may have limited consistent interpretation; undoubtedly, static images of film media were associated with challenges to accurate diagnoses.

Previous multicenter studies have not focused on local reader accuracy of CUS interpretations. Numerous prospective neurodevelopmental outcome studies of preterm infants have demonstrated strong associations of severe CUS abnormalities with neuromotor and neurocognitive deficits (6–8,27,28), but those analyses have been based solely on local reader interpretations. Similarly, reports from multicenter registries have consistently reported CUS data only from local radiologist interpretation (29–31). Our findings thus represent an important and unique opportunity to better understand the validity of the conclusions we reach with respect to those CUS data. The cumbersome logistics of central reading make it unlikely that any further analyses on this scale will be forthcoming. However, our study can offer some reassurance to future investigators using local CUS interpretations. For instance, if a planned study outcome is severe hemorrhagic abnormality, results of local interpretation could be expected to be reasonably accurate. Conversely, sensitivity and specificity of lower grades of IVH and PVL alone in our analysis are disappointing. Although infants with lower grades of IVH have traditionally been considered to have a favorable neurodevelopmental prognosis (3,32,33), a recent study disputes those assumptions. Patra et al found that ELBW infants with grades 1 or 2 IVH had significantly lower Bayley MDI scores, higher rates of major neurologic abnormalities, and higher overall rates of neurodevelopmental impairment at 20 months corrected age than those with no IVH (34). The question remains whether grade 1 or 2 IVH, if accurately and consistently diagnosed, could be a risk factor for later neuromotor impairments. This uncertainty coupled with the observed poor sensitivity for PVL again suggests that imaging techniques such as MRI would be helpful in identifying important patterns of brain injury, notably subtle white matter injury, which could be missed by CUS alone.

There are limitations to this study, some of which are inherent to any retrospective analysis. The data collection instruments for local reader interpretations differed from the central reader data instruments, particularly with respect to the level of detail. Information pertaining to ventriculomegaly, a finding known to be associated with adverse neurodevelopmental outcome (9,10), was not collected on local data instruments. A “rolling” central reading approach, with periodic intraobserver reliability checks, was not part of the PiNO trial because the original scope of the central read was anticipated to be quite small. Furthermore, local reader accuracy analysis was not an aim of the PiNO trial, and the opportunity for a large-scale accuracy analysis was unanticipated; therefore, our findings provide a “real-life” view rather than a “best-case scenario”.

In conclusion, our analysis of CUS data from the PiNO trial demonstrates excellent central reader agreement and local reader accuracy for severe IVH and major prognostic categories. For lower grade IVH and for PVL, reliability and accuracy were poor. These results suggest the validity of reports of major CUS diagnoses, but reports of lower grade IVH and PVL alone should be interpreted with circumspection. These data also reinforce the need to urgently consider expansion of our routine imaging armamentarium to include modalities such as MRI, which would allow for substantially improved discrimination of white matter and other brain injury.

REFERENCES

1. Pape KE, Blackwell RJ, Cusick G, Sherwood A, Houang MT, Thorburn RJ, et al. Ultrasound detection of brain damage in preterm infants. *Lancet* 1979;i:1261–1264. [PubMed: 87726]
2. Slovis TL, Kuhns LR. Real-time sonography of the brain through the anterior fontanelle. *AJR* 1981;136:277–286. [PubMed: 6781247]
3. Shankaran S, Slovis TL, Bedard MP, Poland RL. Sonographic classification of intracranial hemorrhage: a prognostic indicator of mortality, morbidity, and short-term neurologic outcome. *Pediatrics* 1982;100:469–475.

4. Papile L, Burstein J, Burstein R, Koffler H. Incidence and evaluation of subependymal hemorrhage: a study of children with birthweight less than 1500 g. *J Pediatr* 1978;92:529–534. [PubMed: 305471]
5. Volpe JJ. Neurobiology of periventricular leukomalacia of the premature infant. *Pediatr Res* 2001;50:553–562. [PubMed: 11641446]
6. Vohr BR, Wright LL, Poole WK, McDonald SA. Neurodevelopmental outcomes of extremely low birth weight infants <32 weeks gestation between 1993 and 1998. *Pediatrics* 2005;116:635–643. [PubMed: 16143580]
7. Hintz SR, Kendrick DE, Vohr BR, Poole WK, Higgins RD. Changes in neurodevelopmental outcomes at 18–22 months' corrected age among infants of less than 25 weeks gestational age born 1993–1999. *Pediatrics* 2005;115:1645–1651. [PubMed: 15930228]
8. Wood NS, Costeloe K, Gibson AT, Hennessy EM, Wilkinson AR. The EPICure study: associations and antecedents of neurological and developmental disability at 30 months of age following extremely preterm birth. *Arch Dis Child Fetal Neonatal Ed* 2005;90:F134–F140. [PubMed: 15724037]
9. Hack M, Wilson-Costello D, Friedman F, Taylor GH, Schluchter M, Fanaroff AA. Neurodevelopment and predictors of outcomes of children with birth weights of less than 1000 g: 1992–1995. *Arch Pediatr Adolesc Med* 2000;154:725–731. [PubMed: 10891026]
10. Kuban K, Sanocka U, Leviton A, Allred EN, Pagano M, Dammann O, et al. The Developmental Epidemiology Network. White matter disorders of prematurity: association with intraventricular hemorrhage and ventriculomegaly. *J Pediatr* 1999;134:539–546. [PubMed: 10228286]
11. O'Shea RM, Volberg F, Dillard RG. Reliability of interpretation of cranial ultrasound examinations of very low-birthweight neonates. *Dev Med Child Neurol* 1993;35:97–101. [PubMed: 8444332]
12. Pinto J, Paneth N, Kazam E, Kairam R, Wallenstein S, Rose W, et al. Interobserver variability in neonatal cranial ultrasonography. *Paediatric and Perinatal Epidemiology* 1988;2:43–58. [PubMed: 3070483]
13. Corbett SS, Rosenfeld CR, Lupton AR, Risser R, Maravilla AM, Dowling S, et al. Intraobserver and interobserver reliability in assessment of neonatal cranial ultrasounds. *Early Human Development* 1991;27:9–17. [PubMed: 1802667]
14. Graziani LJ, Pasto M, Stanley C, Steben J, Desai H, Desai S, et al. Cranial ultrasound and clinical studies in preterm infants. *J Pediatr* 1985;106:269–276. [PubMed: 3881581]
15. Van Meurs KP, Wright LL, Ehrenkranz RA, Lemons JA, Ball MB, Poole WK, et al. Inhaled nitric oxide for premature infants with severe respiratory failure. *NEJM* 2005;353:13–22. [PubMed: 16000352]
16. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 1977;33:159–174. [PubMed: 843571]
17. Fleiss, JL. *Statistical Methods for Rates and Proportions*. New York: John Wiley; 1981. p. 218
18. Byrt T. How good is that agreement? *Epidemiology* 1996;7:561. [PubMed: 8862998]
19. Altman, DG. *Practical Statistics for Medical Research*. London, England: Chapman and Hall; 1991. p. 403–409.
20. Cicchetti CV, Allison T. A new procedure for assessing reliability of scoring EEG sleep recording. *Am J EEG Tech* 1971;11:101–109.
21. Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ. Psychol. Meas* 1973;33:613–619.
22. Thompson WD, Walter SD. A reappraisal of the kappa coefficient. *J Clin Epidemiol* 1988;41:949–958. [PubMed: 3057117]
23. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol* 1993;46:423–429. [PubMed: 8501467]
24. Cicchetti DV, Feinstein AR. High agreement but low kappa: II Resolving the paradoxes. *J Clin Epidemiol* 1990;43:551–558. [PubMed: 2189948]
25. DeVries LS, van Haastert I-L, Rademaker KJ, Koopman C, Groenendaal F. Ultrasound abnormalities preceding cerebral palsy in high-risk preterm infants. *J Pediatr* 2004;144:815–820. [PubMed: 15192633]
26. Mirmiran M, Barnes PD, Keller K, Constantinou JC, Fleisher BE, Hintz SR, et al. Neonatal brain magnetic resonance imaging before discharge is better than serial cranial ultrasound in predicting

- cerebral palsy in very low birth weight preterm infants. *Pediatrics* 2004;114:992–998. [PubMed: 15466096]
27. The Victorian Infant Collaborative Study Group. Outcome at 2 years of children 23–27 weeks' gestation born in Victoria in 1991–92. *J Paediatr Child Health* 1997;33(2):161–165. [PubMed: 9145362]
 28. Tommiska V, Heinonen K, Ikonen S, Kero P, Pokela M-L, Renlund M, et al. A national short-term follow-up study of extremely low birth weight infants born in Finland in 1996–1997. *Pediatrics* 2001;107:e2. [PubMed: 11134466]
 29. The Victorian Infant Collaborative Study Group (VICS). Improved outcome into the 1990's for infants weighing 500–999g at birth. *Arch Dis Child Neonatal Ed* 1997;77:F91–F94.
 30. Lemons J, Bauer C, Oh W, Korones SB, Papile LA, Stoll BJ, et al. Very low birth weight outcomes of the National Institute of Child Health and Human Development Neonatal Research Network, January 1995 through December 1996. *Pediatrics* 2001;107:e1. [PubMed: 11134465]
 31. Horbar JD, Badger GJ, Carpenter JH, Fanaroff AA, Kilpatrick S, LaCorte M, et al. Trends in mortality and morbidity for very low birth weight infants, 1991–1999. *Pediatrics* 2002;110:143–151. [PubMed: 12093960]
 32. Fawer C-L, Calame A, Furrer M-T. Neurodevelopmental outcome at 12 months of age related to cerebral ultrasound appearance of high risk preterm infants. *Early Human Dev* 1985;11:123–132.
 33. Vaucher, YE.; Perritt, R.; Finer, NN.; Higgins, RD. the NICHD Neonatal Research Network. Uncomplicated grades 1 and 2 intraventricular hemorrhage are not associated with early childhood neurodevelopmental disability in ELBW infants. Washington, DC: Presented to the Society for Pediatric Research; 2005 May 14–17.

APPENDIX

The Preemie Inhaled Nitric Oxide Study Group

Brown University Women & Infant's Hospital Principal Investigator: William Oh, MD; Study Coordinator Angelita Hensman, BSN, RNC; Respiratory Therapist: Daniel Gingras, RRT.

Emory University Principal Investigators: Barbara J. Stoll, MD and Lucky Jain, MD; Study Coordinator: Ellen Hale, RN, BS; Respiratory Therapist: Irma Seabrook, BS, RRT-NPS.

Indiana University Riley Hospital for Children and Methodist Hospital Principal Investigators Greg Sokol, MD and Dianne Lorant, MD; Study Coordinators: Diana Dawn Appel, RN BSN and Lucy Miller, RN BSN; Respiratory Therapists: Dale Chriscinske, BS, RRT, NPS and Jeff Attwood, RRT.

Northwestern University Principal Investigator Robin Steinhorn, MD; Study Coordinator and Respiratory Therapist: Michael Sautel, RRT.

Stanford University Principal Investigator: Krisa VanMeurs, MD; Study Coordinator Bethany Ball, BS, CCRC; Respiratory Therapist: Dan Proud, RCP.

University of Alabama at Birmingham University Hospital-UAB Principal Investigator: Waldemar A. Carlo, MD; Study Coordinator: Shirley S. Cosby, RN, BSN; Respiratory Therapist: Robert B. Johnson RRT.

University of Cincinnati University Hospital, Cincinnati Children's Hospital Medical Center and Good Samaritan Principal Investigators Jon Fridriksson, MD and Barb Warner MD; Study Coordinators: Marcia Mersmann, RN, Barb Alexander, RN, Jody Shively, RN, Holly Mincey, RN; Respiratory Therapists: Mary Hoover, RRT, Sharon Sapientz, RRT, Eric Stephenson, RRT.

University of California-San Diego UCSD Medical Center and Sharp Mary Birch Hospital for Women Principal Investigators: Neil N. Finer, MD and Maynard R. Rasmussen, MD; Study Coordinators: Chris Henderson, CRTT and Clarence Demetrio, RN; Respiratory Therapists: Wade Rich, RRT-NPS and Christine Joseph, RRT-NPS.

University of Florida Wolfson Children's Hospital at Baptist Medical Center and Shands Jacksonville Medical Center Principal Investigator: Mark Hudak, MD; Study Coordinators: Shannon Osbeck, RN, BSN and Elizabeth Case, RN, BSN, CCRC; Respiratory Therapists: Amanda Kellum, RRT and Lamont Hogans, RRT.

University of Rochester Golisano Children's Hospital at Strong Principal Investigator: Carl T. D'Angio, MD; Study Coordinator: Linda Reubens, RN; Respiratory Therapist: Greg Hutton, RRT.

University of Texas – Dallas Parkland Hospital Principal

Investigator: Abbot Laptook, MD; Study Coordinators: Susie Madison, RN, Gay Hensley, RN and Nancy Miller, RN; Respiratory Therapist: Glenn Metoyer, RRT. **University of Texas – Houston** Memorial Hermann Children’s Hospital Principal Investigator: Kathleen Kennedy, MD, MPH; Study Coordinator: Georgia McDavid, RN; Respiratory Therapist: Danny Emerson, BA, RRT, RCP. **Medical College of Wisconsin** Principal Investigator: Ganesh Konduri, MD; Study Coordinator: Mike Paquette, RCP/CRT; Respiratory Therapists: Steven Wong, Mike Paquette, RCP/CRT. **Wake Forest University** Wake Forest University Baptist Medical Center, Forsyth Medical Center and Brenner Children’s Hospital Principal Investigators: Judy Aschner, MD and T. Michael O’Shea, MD, MPH; Study Coordinator: Nancy Peters, RN and B.J. Hansell, RRT, CCRC; Respiratory Therapists: Jennifer Griffin, RRT, RCP and Clay Adams, RRT, RCP. **Wayne State University** Hutzel Women's Hospital & Children's Hospital of Michigan Principal Investigator: Seetha Shankaran, MD; Study Coordinators: Rebecca Bara, RN, BSN and Geraldine Muran, RN, BSN; Respiratory Therapist: Wonder Weekfall, RRT. **Yale University** New Haven Children's Hospital Principal Investigator: Richard A. Ehrenkranz, M.D. Study Coordinator: Patricia Gettner, RN; Respiratory Therapist: Art Caldwell, AS, RRT. **RTI Central Reading staff** Deborah Schwartz; Kim Doeden; Carolyn Petrie; Neha Patel

Appendix

Appendix II
NICHD Neonatal Research Network grants 1996–2006

PI	Center, Location	NICHD Grant #	GCRC #
Waldemar A. Carlo, MD	University of Alabama at Birmingham Birmingham, Alabama	U10 HD34216	
Edward F. Donovan, MD	University of Cincinnati Cincinnati, Ohio	U10 HD27853	M01 RR 08084
Richard A. Ehrenkranz, MD	Yale University New Haven, Connecticut	U10 HD27871	M01 RR 06022
Neil N. Finer, MD	University of California at San Diego San Diego, CA	U10 HD40461	
Abbot R. Laptook, MD	University of Texas Southwestern Medical Center at Dallas Dallas, Texas	U10 HD40689	
James A. Lemons, MD	Indiana University Indianapolis, Indiana	U10 HD27856	M01 RR 00750
William Oh, MD	Women and Infants’ Hospital of Rhode Island Providence, Rhode Island	U10 HD27904	
T. Michael O’Shea, MD	Wake Forest University School of Medicine Winston-Salem, NC	U10 HD40498	
Dale L. Phelps, MD	University of Rochester Rochester, New York	U10 HD40521	5 M01 RR00044
W. Kenneth Poole, PhD	Research Triangle Institute	U01 HD36790	
Seetha Shankaran, MD	Wayne State University Detroit, MI	U10 HD21385	
David K. Stevenson, MD	Stanford University Stanford, CA	U10 HD27880	M01 RR 00070
Barbara J. Stoll, MD	Emory University Atlanta, GA	U10 HD27851	
Jon E. Tyson, MD, MPH	The University of Texas Health Science Center at Houston	U10 HD 21373	

PI	Center, Location	NICHD Grant #	GCRC #
	Houston, TX		
Rosemary D. Higgins, MD	National Institute of Child Health and Human Development Bethesda, MD		

Center:		Study No: P	Network No:	Birth No	DATE April 7, 2004
A. IDENTIFICATION					
1. READER:	INITIALS ____/____/____			5. QUALITY:	GOOD POOR
2. DATE READ:	MONTH DAY YEAR ____/____/____			6. READABLE:	YES NO
3. READING:				8. ALL NECESSARY VIEWS AVAILABLE:	YES NO
4. DATE OF SONOGRAM:				9. NORMAL READING:	YES NO

ULTRASOUND RESULTS	Left		Right	
	Yes	No	Yes	No
B. ECHODENSITY SITE				
1. Echodensity present				
If Yes, mark all that apply				
a. Subependymal				
b. Periventricular (PVL)				
If Yes, mark all that apply				
1. Frontal				
2. Temporal				
3. Parietal				
4. Occipital				
c. Intraventricular				
1) If Yes,				
< 25 % filled				
25-50 % filled				
> 50% filled				
d. Intracerebral				
If Yes, mark all that apply				
1. Frontal				
2. Temporal				
3. Parietal				
4. Occipital				
5. Thalamus				
6. Posterior Fossa				
e. Other				

	Left		Right	
	Yes	No	Yes	No
C. ECHOLUCENCY SITE				
1. Echolucency present				
If Yes, mark all that apply				
a. Subependymal				
b. Periventricular (PVL)				
If Yes, mark all that apply				
1. Frontal				
2. Temporal				
3. Parietal				
4. Occipital				
c. Intracerebral				
If Yes, mark all that apply				
1. Frontal				
2. Temporal				
3. Parietal				
4. Occipital				
d. Parenchymal Cyst				
e. Other				
D. OTHER PARENCHYMAL				
1. Other Parenchymal				
If Yes, mark all that apply				
a. Calcification				
b. Cerebral edema				
c. Cortical atrophy				
d. Extra axial fluid				
e. Infarct				
f. LS Branching				
g. Other				

	Left		Right	
	Yes	No	Yes	No
E. VENTRICLES				
1. Ventricle abnormality				
a. If Yes,				
Mild increase				
Moderate increase				
Severe increase				
Slit Ventricles				
2. Choroid abnormality				
If Yes, mark all that apply				
a. Cyst				
b. Hemorrhage				
c. other				
F. HEMORRHAGE CLASSIFICATION				
1. Hemorrhage				
If Yes,				
a. Papile Classification				
Grade I				
Grade II				
Grade III				
Grade IV				
Indeterminate				
G. STRUCTURAL ABNORMALITIES				
1. Structural abnormality				
If Yes, comments:				

Figure 1.
Central Reader data instrument from the NICHD Neonatal Research Network PiNO Trial

Table 1

Central reader diagnostic assignments. Data shown as N (%)

	Central Reader #1	Central Reader #2
Normal *	177 (54%)	133 (41%)
Grade 1	19 (3%)	81 (13%)
Grade 2	35 (5%)	4 (1%)
Grade 3	84 (13%)	103 (16%)
Grade 4	74 (11%)	62 (10%)
Grade 1 OR 2	54 (8%)	85 (13%)
Grade 3 OR 4	158 (24%)	165 (26%)
Echodense PVL	3 (0%)	7 (1%)
Echolucent PVL	8 (1%)	1 (0%)
Any ventricular enlargement	195 (30%)	246 (38%)
Grade 3 or 4 or PVL (any)	165 (25%)	171 (27%)

* Total N=326 for “Normal” reading; other diagnostic categories were based on hemispheres

Table 2

Interobserver reliability of CR #1 and CR #2 for major diagnostic and prognostic categories

	N [*]	Kappa	95%CI	PPA [†]
Normal vs. Abnormal	326	0.60	0.52 – 0.68	78.7
IVH grade 1	644	0.26	0.15 – 0.38	30.0
IVH grade 2	644	0.20	0.03 – 0.36	20.5
IVH grade 3	644	0.64	0.55 – 0.72	68.8
IVH grade 4	644	0.66	0.56 – 0.76	69.6
IVH grade 1 OR 2	644	0.40	0.29 – 0.51	46.0
IVH grade 3 OR 4	644	0.84	0.79 – 0.89	87.9
Any PVL ^{**}	642	0.09	–0.10 – 0.28	10.5
Grade 3 or 4 or PVL	641	0.81	0.75 – 0.86	85.6
Any ventricular increase	643	0.76	0.71 – 0.81	84.2

* Normal vs. abnormal was a global assessment, whereas other analyses were hemisphere-specific; observations for all categories were not available due to missing or uninterpretable data.

[†] PPA=percent positive agreement

** Either echolucent or echodense PVL

Table 3
Reliability table of central reader interpretations: IVH grade*

		CR #2					
		No IVH	Grade 1	Grade 2	Grade 3	Grade 4	Total
CR #1	No IVH	372	44	0	7	3	426
	Grade 1	4	15	0	0	0	19
	Grade 2	9	13	4	9	0	35
	Grade 3	3	5	0	64	11	83
	Grade 4	3	4	0	19	47	73
	Total	391	81	4	99	61	636

* CR #1=Central Reader #1; CR #2=Central Reader #2

Weighted kappa, (95%CI) = 0.79, (0.75 – 0.82)

Table 4
Reliability table of central reader interpretations: ventricular enlargement*

CR #2		None	Mild	Moderate	Severe	Total
CR #1	None	390	49	11	0	450
	Mild	7	28	29	2	66
	Moderate	2	12	47	27	88
	Severe	0	0	0	39	39
	Total	399	89	87	68	643

* CR #1=Central Reader #1; CR #2=Central Reader #2

Weighted kappa, (95%CI) = 0.75, (0.71 – 0.79)

Table 5

Sensitivities and specificities for local interpretations of major diagnostic and prognostic categories compared with central readers

	CENTRAL READER #1		CENTRAL READER #2	
	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)
Any IVH	92.2	86.4	84.2	87.7
Grade 1	53.3	95.1	28.0	96.7
Grade 2	47.4	95.4	50.0	93.2
Grade 3	61.5	94.2	41.9	94.7
Grade 4	82.3	95.7	86.0	92.0
Grade 1 OR 2	67.7	92.1	48.1	92.3
Grade 3 OR 4	92.0	92.4	88.2	93.1
PVL	44.4	98.1	20.0	97.2
Grade 3 or 4 or PVL	90.3	93.4	86.5	91.6

Table 6
Grade of IVH – local interpretations compared with central readers

Central Reader #1						
	No IVH	Grade 1	Grade 2	Grade 3	Grade 4	Total
No IVH	171	3	3	2	0	179
Grade 1	9	8	3	1	2	23
Grade 2	9	3	9	1	1	23
Grade 3	4	1	4	16	8	33
Grade 4	5	0	0	6	51	62
Total	198	15	19	26	62	320

LOCAL

Central Reader #2						
	No IVH	Grade 1	Grade 2	Grade 3	Grade 4	Total
No IVH	157	18	0	5	0	180
Grade 1	7	14	0	0	2	23
Grade 2	8	10	1	4	0	23
Grade 3	5	4	1	18	5	33
Grade 4	2	4	0	16	43	65
Total	179	50	2	43	50	324

LOCAL

Table 7

Presence of PVL – local interpretations compared with central readers

Central Reader #1				
	Any PVL	No PVL	Total	LOCAL
Any PVL	4	6	10	
No PVL	5	310	315	
Total	9	316	325	

Central Reader #2				
	Any PVL	No PVL	Total	LOCAL
Any PVL	1	9	10	
No PVL	4	311	315	
Total	5	320	325	

Table 8

Grade 3 or 4 IVH or PVL: Local compared interpretations with central readers

Central Reader #1				
	Grade 3 or 4 or PVL	No grade 3 or 4 or PVL	Total	
Grade 3 or 4 or PVL	84	15	99	LOCAL
No grade 3 or 4 or PVL	9	211	220	
Total	93	226	319	

Central Reader #2				
	Grade 3 or 4 or PVL	No grade 3 or 4 or PVL	Total	
Grade 3 or 4 or PVL	83	19	102	LOCAL
No grade 3 or 4 or PVL	13	208	221	
Total	96	227	323	