



Published in final edited form as:

*Pediatr Radiol.* 2007 December ; 37(12): 1201–1208. doi:10.1007/s00247-007-0605-z.

## Observer variability assessing US scans of the preterm brain: the ELGAN study

**Karl Kuban,**

Division of Pediatric Neurology, Boston University Medical Center, Boston University School of Medicine, Boston, MA, USA

**Ira Adler,**

Eastern Radiologists, Inc., Greenville, NC, USA

**Elizabeth N. Allred,**

Neuroepidemiology Unit, Children's Hospital Boston, Harvard Medical School, Harvard School of Public Health, Boston, MA, USA

**Daniel Batton,**

Departments of Pediatrics and Neonatology, William Beaumont Hospital, Royal Oak, MI, USA

**Steven Bezinque,**

Department of Radiology, DeVos Children's Hospital, Grand Rapids, MI, USA

**Bradford W. Betz,**

Department of Radiology, DeVos Children's Hospital, Grand Rapids, MI, USA

**Ellen Cavenagh,**

Department of Radiology, Sparrow Hospital, Lansing, MI, USA

**Sara Durfee,**

Department of Radiology, Brigham & Women's Hospital, Harvard Medical School, Boston, MA, USA

**Kirsten Ecklund,**

Department of Radiology, Children's Hospital Boston, Harvard Medical School, Boston, MA, USA

**Kate Feinstein,**

Department of Radiology, University of Chicago Hospital, University of Chicago, Chicago, IL, USA

**Lynn Ansley Fordham,**

Department of Radiology, University of North Carolina School of Medicine, Chapel Hill, NC, USA

**Frederick Hampf,**

Department of Radiology, Baystate Medical Center, Springfield, MA, USA

**Joseph Junewick,**

Department of Radiology, DeVos Children's Hospital, Grand Rapids, MI, USA

**Robert Lorenzo,**

Department of Radiology, Children's Healthcare of Atlanta, Emory University School of Medicine, Atlanta, GA, USA

**Roy McCauley,**

Department of Radiology, Tufts-New England Medical Center, Tufts University School of Medicine, Boston, MA, USA

**Cindy Miller,**

Department of Radiology, Yale-New Haven Hospital, Yale University School of Medicine, New Haven, CT, USA

**Joanna Seibert,**

Department of Radiology, Arkansas Children's Hospital, University of Arkansas Medical School, Little Rock, AR, USA

**Barbara Specter,**

Department of Radiology, Forsyth Hospital, Baptist Medical Center, Wake Forest University School of Medicine, Winston-Salem, NC, USA

**Jacqueline Wellman,**

Department of Radiology, Milford Regional Medical Center, Milford, MA, USA

**Sjirk Westra,** and

Division of Pediatric Radiology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

**Alan Leviton**

Neuroepidemiology Unit, Children's Hospital Boston, Harvard Medical School, 1 Autumn St. #720, Boston, MA 02215-5393, USA, alan.leviton@childrens.harvard.edu

**Abstract**

**Background**—Neurosonography can assist clinicians and can provide researchers with documentation of brain lesions. Unfortunately, we know little about the reliability of sonographically derived diagnoses.

**Objective**—We sought to evaluate observer variability among experienced neurosonologists.

**Materials and methods**—We collected all protocol US scans of 1,450 infants born before the 28th postmenstrual week. Each set of scans was read by two independent sonologists for the presence of intraventricular hemorrhage (IVH) and moderate/severe ventriculomegaly, as well as hyperechoic and hypoechoic lesions in the cerebral white matter. Scans read discordantly for any of these four characteristics were sent to a tie-breaking third sonologist.

**Results**—Ventriculomegaly, hypoechoic lesions and IVH had similar rates of positive agreement (68–76%), negative agreement (92–97%), and kappa values (0.62 to 0.68). Hyperechoic lesions, however, had considerably lower values of positive agreement (48%), negative agreement (84%), and kappa (0.32). No sonologist identified all abnormalities more or less often than his/her peers. Approximately 40% of the time, the tie-breaking reader agreed with the reader who identified IVH, ventriculomegaly, or a hypoechoic lesion in the white matter. Only about 25% of the time did the third party agree with the reader who reported a white matter hyperechoic lesion.

**Conclusion**—Obtaining concordance seems to be an acceptable way to assure reasonably high-quality of images needed for clinical research.

**Keywords**

Brain; Newborn; Premature

**Introduction**

Imaging in medicine requires interplay between the imaged subject and the viewer. The intent, most often, is to maximize resolution and minimize the contribution of observer variability. Yet, physicians sometimes interpret images differently, including images generated from US

studies in adults [1–4] and newborns [5–9]. This has obvious clinical as well as research implications [10], and may affect the value of studies that seek to identify the correlates and antecedents of US scan images.

We are aware of only a small number of studies that have evaluated how reliably experienced sonologists read head US scans of infants [5–9]. Except for one study [9], only a small proportion of infants in any of these studies was born before 28 weeks, when the prevalence of abnormalities is highest.

Our ELGAN study, so-called because all subjects were *extremely low gestational age newborns* (ELGANs), required all protocol US scans of 1,450 infants to be read separately by two independent readers. Each reader was asked to identify the four most common cranial sonographic findings associated with subsequent neurological impairment in ELGANs: intraventricular hemorrhage (IVH), ventricular enlargement, white matter hyperechoic lesions, and white matter hypoechoic lesions. Scans not read entirely concordantly by the first two readers were sent to another independent reader for a tie-breaking interpretation. These structural aspects of the ELGAN study provided the opportunity to assess observer variability among 14 highly experienced sonologists who each read more than 200 sets of scans also read by others.

## Materials and methods

### Sample

The sample for this study consisted of all babies born between 23 and 27 completed weeks of gestation at 14 hospitals in 11 cities in five states between March 2002 and August 2004. The babies had at least one cranial US scan, and their parents consented to their participation. The Institutional Review Board at each participating institution approved this study.

### Protocol scans

Routine scans were performed by sonographers at all of the hospitals using high-frequency transducers (7.5 and 10 MHz), most often with Acuson Sequoia, Philips iU55, and GE Logiq9 systems. US studies always included the six standard quasicoronal views and five para-sagittal views using the anterior fontanel as the sonographic window [11].

A total of 1,450 infants had at least one set of US scans, and 895 had three sets. A set of scans was obtained in 1,123 infants between the 1st and 4th days of life, in 1,302 between the 5th and 14th days of life, and in 1,268 between the 15th day of life and 40 weeks postmenstrual age.

### Training

All sonologists reviewed a manual and data collection form used in a previous study [12–14], and some suggested revisions. Subsequently, all agreed to the revised manual and data collection form. The manual and data collection form included templates of multiple levels of ventriculomegaly. The cerebral white matter in each hemisphere was divided into eight zones chosen for ease of identification by ultrasonographic landmarks. The lesions in each zone could be further characterized as hyperechoic and/or hypoechoic.

The need for additional revision was assessed with scans that illustrated a variety of abnormalities that could be interpreted differently by experienced sonologists. These scans were distributed and served as discussion points for conference calls intended to work out ways to minimize variation in interpretation. Potential sources of variation were identified, most of which had already been identified and addressed in the manual.

Intraventricular hemorrhage (IVH) was defined as blood within the ventricles. By definition, this excluded hemorrhage localized to the subependymal region. Ventriculomegaly, categorized as mild, moderate, and severe, was defined visually with a template that was on the data collection form. Our emphasis was on moderate/severe ventriculomegaly, which was diagnosed if the lateral ventricle was at least moderately enlarged in any of four sections (frontal horn, body, and occipital horn) on either side.

### Reading procedures

All US scans were read by two independent readers who were not provided with clinical information. One reader was the study sonologist at the infant's birth institution who for this study read all the protocol scans for each baby at one time. The other was the study sonologist at another ELGAN study institution who was provided with the scans as electronic images on a CD imbedded in the software eFilm Workstation (Merge Healthcare, Milwaukee, Wis.). The eFilm program allowed the second reader to see the images as the first reader had seen them, and to adjust and enhance the images, including the ability to zoom and alter contrast and brightness.

Definitions of discrepancies that required resolution by a third (tie-breaking) reader included:

|                    |   |
|--------------------|---|
| IVH                | One reader classified an infant's scans as having no or an uncertain IVH and the other reader classified the scan as having probable or definite IVH.                 |
| Ventriculomegaly   | One reader classified an infant's scans as having no or mild ventriculomegaly and the other reader classified the scan as having moderate or severe ventriculomegaly. |
| Hyperechoic lesion | One reader classified an infant's scans as having no hyperechoic lesion in the white matter and the other reader reported the presence of a hyperechoic lesion.       |
| Hypoechoic lesion  | One reader classified an infant's scans as having no hypoechoic lesion in the white matter and the other reader reported the presence of a hypoechoic lesion.         |

Approximately 40% of scan sets were not read entirely concordantly by the first two readers. These were sent to a third reader, who was not informed of the nature of the discrepancy and was asked to complete the entire data collection form for all of the infant's protocol scans.

### Data analysis

This descriptive study characterizes the more common discrepancies seen when cranial US scans were read by multiple readers. Kappa values are used as a statistical descriptor of observer variability [15], but we recognize their limitations [16,17].

We evaluated the following hypotheses:

1. Inter-reader reliability differs for the four major cranial US findings; identification of a hyperechoic lesion is the least reliable and identification of IVH the most reliable.
2. Familiarity with an institution's equipment and procedures influences the propensity to identify an abnormality.
3. The third reader more often confirms the presence of an abnormality than its absence.
4. Agreement will be higher for studies performed closer to term than those performed in the first week or two of life.

## Results

### Hypothesis 1: Some US images are identified more reliably than others

According to congruent readings, the overall prevalence of the four target findings were: IVH, 24%; ventriculomegaly, 12%; hyperechoic lesion, 16%; and hypoechoic lesion, 7.8% (Table 1). Ventriculomegaly, echolucency, and IVH tended to be grouped together in their positive

agreement rates (68–76%), negative agreement rates (92–97%), and kappa values (0.62 to 0.68; Table 2). In contrast, hyperechoic lesions had considerably lower values of positive agreement (48%), negative agreement (84%), and kappa (0.32).

A total of 1,450 sets of scans were read separately by two readers. Of all these scans, 12% were read discordantly for IVH, 9% for moderate/severe ventriculomegaly, 24% for hyperechoic lesion and 6% for hypoechoic lesion.

We compared how each reader fared in relation to all other second readers who read the same scans in an attempt to identify whether particular readers stood out as consistently identifying abnormalities more or less often than others (Table 3). Several patterns emerged from these analyses.

First, although certain individual readers identified a lesion more or less frequently than colleagues, most often this tendency was lesion-specific. For example, reader H identified IVH a bit more often than others who read the same sets of scans, but reader H did not identify ventriculomegaly or white matter lesions more frequently. Second, the identification of a white matter hyperechoic lesion showed the most variability (Table 3). One reader recognized a hyperechoic finding in almost half and five other readers identified a hyperechoic lesion in approximately 20% of scans that another colleague interpreted as having no hyperechoic lesion. Eight readers differed from their colleagues in their recognition of a hyperechoic lesion in at least 25% of the scan sets. Third, a hypoechoic lesion had the highest agreement rates, even higher than those for intraventricular hemorrhage. Fourth, the choice of discrepancy definition strongly influenced the concordance rate for ventriculomegaly. When readings were allocated to one of two categories, either normal (combining all normal or mildly abnormal readings) or abnormal (combining all moderately and severely abnormal readings), 9% of readings were discordant for ventriculomegaly. When categories were not collapsed into two groups and readings were viewed as concordant if they were within one category or the other, the rate of discordance was only 3%. In analyses that maintained four categories of ventriculomegaly, the larger the ventriculomegaly the higher the association with the other three sonographic findings. This was most evident for hyperechoic and hypoechoic lesions, with the risk of these characteristics doubling with each increase in size of ventriculomegaly (Table 4).

### **Hypothesis 2: Third readers more often align with the reader from an outside center rather than with the “home institution” reader**

To address whether familiarity with an institution’s equipment influenced readings, we compared concordance rates between the second and third readers and the first and third readers. The data from readings for ventriculomegaly and for a hyperechoic lesion supported this view of greater similarity between the second and third readers (Table 5). On the other hand, for intraventricular hemorrhage and hypoechoic lesion, the agreement between first and third readers was very similar to the agreement between second and third readers.

### **Hypothesis 3: The third reader more often confirms the presence than the absence of an abnormality**

For intraventricular hemorrhage, moderate/severe ventriculomegaly, and a hypoechoic lesion, the third reader read approximately 40% of discrepant scans as positive, which clearly indicates a tendency to identify these lesions less often than earlier readers. This tendency was even more pronounced for a hyperechoic lesion. Third readers identified a hyperechoic lesion on 19% of 165 scans read by first readers and on 26% of 183 scans read by second readers (23% of scans overall; Table 5).

#### **Hypothesis 4: Reliability is higher for studies performed closer to term than for those performed in the first week or two of life**

In this analysis, we compared the rates of discrepancy between first and second readers separately for each set of protocol scans. The first two protocol scans were obtained in the first 2 weeks of life, while the third protocol scan was obtained closer to term (Table 6). A hyperechoic lesion was the only finding for which agreement improved from early to late scans, but even this improvement in agreement was modest.

### **Discussion**

Cranial US readings in this cohort of 1,450 ELGANs were most reliable for ventriculomegaly, a hypoechoic lesion, and IVH, and least reliable for an echogenic lesion. The kappa value for a hyperechoic lesion, an index of the overall level of reliability that corrects for the level of agreement that would be expected by chance alone, was half that for the other findings, which ranged from 0.62 to 0.68. Some readers thought their assessments were limited by not having additional (posterior fontanel and trans-mastoid) views that were obtained routinely at their institution, by not having cine clips, and by not being able to scan the infant themselves. Nevertheless, the reading discrepancies among the sonologists in our study could not be explained by whether the studies were read by a reader from the birth hospital or from an outside collaborating institution, nor whether the scans were performed closer to term or in the first week or two of life.

Cranial US variability has been investigated by at least four groups [6–9]. The positive and negative agreement rates for IVH in this study are comparable to those of one study [7] and modestly lower than those of the three others [6,8,9]. In one of the previous studies, “echolucent or echodense periventricular leukomalacia” had a kappa of 0.09 [9]. In another study, in which kappa values for periventricular leukomalacia were found to range from 0.17 to 0.22, the “gold standard” outside reviewers found periventricular leukomalacia to be present five times more often than the initial home-center readers [18]. Unlike two studies that had both local and central readers [9,18], local readers in our study functioned as both outside readers for scans obtained at other institutions as well as local readers for scans obtained at their own institutions. Perhaps that is why the prevalence of findings seen by the second, outside readers did not differ substantially from that of the first. All four US findings were defined as either present or absent for the purpose of requiring a third reading. This simplicity minimizes variability when staging and grading schemes are offered [19]. We did, however, explore finer gradations of ventriculomegaly.

Defining differences as any nonadjacent category reduced the variability. We advise caution in drawing inferences based on this finding, in large part because we did not seek third readings based on nonadjacent discrepancies.

This is the largest study in which interobserver variability in the reading of cranial US studies of infants born before the 28th postmenstrual week has been evaluated. The rates of agreement and variability found in our study are likely to be generalizable. They are comparable to those reported for mammograms [20,21], cranial CT scans in assessments of subarachnoid blood [22], cerebral angiograms for determining the completeness of aneurysm therapy [23], CT scans for assessing patients for intravenous rtPA therapy [24], and MRI looking at white matter changes in adults [25].

In a previous study we used a consensus approach to the reading of cranial US scans of very low birth weight infants [13,14]. Three sonologists reading together were able to agree on the presence or absence of all abnormalities on 90% of scans sent to them because a screening sonologist identified an abnormality. Others, too, have found that reliability is enhanced by

consensus readings [26,27]. Although we prefer a consensus approach, we found it to be not feasible for a large number of sonologists to come together to review more than 500 sets of studies. In addition, at that time we did not have the capability to share all images comparably during conference calls. With the availability of CDs and visualization techniques identical to those available at the home institution, conference calls among three or more readers are certainly possible, and this is the approach we recommend.

Efforts are underway to improve the reading of scans. Some have employed scales to assess how confident physicians are in reading fetal MRI scans [28]. Evaluating the correlates of confidence might help identify characteristics of scans or readers that contribute to observer variability. Still others are trying to improve the quality of cranial US images [29]. The higher the quality of the data, the more likely an epidemiologic study will provide useful information [10]. The quality of US scans can be assessed against a gold standard, such as neuropathology or MRI. Until early and repeated MRI scans are obtained routinely, however, cranial sonograms will continue to be used for clinical and research purposes.

Although variability in cranial US readings is higher than we would like, seeking concordance (i.e. at least two readers must agree) seems to be an acceptable way to assure a reasonably high quality of interpretation of images needed for clinical research. Observer variability can also be reduced by additional training designed specifically to reduce interreader disparities, or by having all scans read simultaneously by several readers working to achieve consensus.

## Conclusions

We have shown that IVH, moderate/severe ventriculomegaly and a hypoechoic lesion in the white matter can be assessed with acceptable levels of observer agreement. On the other hand, hyperechoic lesions pose a problem. Additional work is needed to see how best to reduce observer variability associated with hyperechoic lesions.

## Acknowledgments

This work was funded by a cooperative agreement with the National Institute of Neurological Disorders and Stroke (1 U01 NS 40069-01A2) and a program project grant from the National Institute of Child Health and Human Development (NIH-P30-HD-18655). The authors are also grateful for the assistance of all their colleagues, and the cooperation of the families of the infants who are the focus of our attention.

## References

1. Griffiths GD, Razzaq R, Farrell A, et al. Variability in measurement of internal carotid artery stenosis by arch angiography and duplex ultrasonography – time for a reappraisal? *Eur J Vasc Endovasc Surg* 2001;21:130–136. [PubMed: 11237785]
2. Ballantyne SA, O’Neill G, Hamilton R, et al. Observer variation in the sonographic measurement of optic nerve sheath diameter in normal adults. *Eur J Ultrasound* 2002;15:145–149. [PubMed: 12423741]
3. Winkfield B, Aube C, Burtin P, et al. Inter-observer and intra-observer variability in hepatology. *Eur J Gastroenterol Hepatol* 2003;15:959–966. [PubMed: 12923367]
4. Berg WA, Blume JD, Cormack JB, et al. Operator dependence of physician-performed whole-breast US: lesion detection and characterization. *Radiology* 2006;241:355–365. [PubMed: 17057064]
5. Pinto J, Paneth N, Kazam E, et al. Interobserver variability in neonatal cranial ultrasonography. *Paediatr Perinat Epidemiol* 1988;2:43–58. [PubMed: 3070483]
6. Pinto-Martin J, Paneth N, Witomski T, et al. The central New Jersey neonatal brain haemorrhage study: design of the study and reliability of ultrasound diagnosis. *Paediatr Perinat Epidemiol* 1992;6:273–284. [PubMed: 1584728]

7. Harris DL, Teele RL, Bloomfield FH, et al. Does variation in interpretation of ultrasonograms account for the variation in incidence of germinal matrix/intraventricular haemorrhage between newborn intensive care units in New Zealand? *Arch Dis Child Fetal Neonatal Ed* 2005;90:F494–F499. [PubMed: 16244209]
8. O'Shea TM, 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]
9. Hintz SR, Slovis T, Bulas D, et al. Interobserver reliability and accuracy of cranial ultrasound scanning interpretation in premature infants. *J Pediatr* 2007;150:592–596. [PubMed: 17517240]
10. Tooth L, Ware R, Bain C, et al. Quality of reporting of observational longitudinal research. *Am J Epidemiol* 2005;161:280–288. [PubMed: 15671260]
11. Teele, R.; Share, J. *Ultrasonography of infants and children*. Philadelphia: Saunders; 1991.
12. Leviton A, Paneth N, Reuss ML, et al. Developmental Epidemiology Network Investigators. Maternal infection, fetal inflammatory response, and brain damage in very low birth weight infants. *Pediatr Res* 1999;46:566–575. [PubMed: 10541320]
13. Kuban K, Sanocka U, Leviton A, 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]
14. Kuban KC, Allred EN, Dammann O, et al. Topography of cerebral white-matter disease of prematurity studied prospectively in 1607 very-low-birthweight infants. *J Child Neurol* 2001;16:401–408. [PubMed: 11417604]
15. Fleiss JL, Kingman A. Statistical management of data in clinical research. *Crit Rev Oral Biol Med* 1990;1:55–66. [PubMed: 2101261]
16. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol* 1993;46:423–429. [PubMed: 8501467]
17. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med* 2005;37:360–363. [PubMed: 15883903]
18. Harris DL, Bloomfield FH, Teele RL, et al. Variable interpretation of ultrasonograms may contribute to variation in the reported incidence of white matter damage between newborn intensive care units in New Zealand. *Arch Dis Child Fetal Neonatal Ed* 2006;91:F11–F16. [PubMed: 16159954]
19. Redline RW, Faye-Petersen O, Heller D, et al. Amniotic infection syndrome: nosology and reproducibility of placental reaction patterns. *Pediatr Dev Pathol* 2003;6:435–448. [PubMed: 14708737]
20. Beam CA, Layde PM, Sullivan DC. Variability in the interpretation of screening mammograms by US radiologists. Findings from a national sample. *Arch Intern Med* 1996;156:209–213. [PubMed: 8546556]
21. Elmore JG, Wells CK, Lee CH, et al. Variability in radiologists' interpretations of mammograms. *N Engl J Med* 1994;331:1493–1499. [PubMed: 7969300]
22. Svensson E, Starmark JE, Ekholm S, et al. Analysis of interobserver disagreement in the assessment of subarachnoid blood and acute hydrocephalus on CT scans. *Neurol Res* 1996;18:487–494. [PubMed: 8985947]
23. Cloft HJ, Kaufmann T, Kallmes DF. Observer agreement in the assessment of endovascular aneurysm therapy and aneurysm recurrence. *AJNR* 2007;28:497–500. [PubMed: 17353321]
24. Grotta JC, Chiu D, Lu M, et al. Agreement and variability in the interpretation of early CT changes in stroke patients qualifying for intravenous rtPA therapy. *Stroke* 1999;30:1528–1533. [PubMed: 10436095]
25. Kapeller P, Barber R, Vermeulen RJ, et al. Visual rating of age-related white matter changes on magnetic resonance imaging: scale comparison, interrater agreement, and correlations with quantitative measurements. *Stroke* 2003;34:441–445. [PubMed: 12574557]
26. de Vet HC, Koudstaal J, Kwee WS, et al. Efforts to improve interobserver agreement in histopathological grading. *J Clin Epidemiol* 1995;48:869–873. [PubMed: 7782794]
27. Kujan O, Khattab A, Oliver RJ, et al. Why oral histopathology suffers inter-observer variability on grading oral epithelial dysplasia: an attempt to understand the sources of variation. *Oral Oncol* 2007;43:224–231. [PubMed: 16931119]



28. Breeze AC, Cross JJ, Hackett GA, et al. Use of a confidence scale in reporting postmortem fetal magnetic resonance imaging. *Ultrasound Obstet Gynecol* 2006;28:918–924. [PubMed: 17124693]
29. Vansteenkiste E, Pizurica A, Philips W. Improved segmentation of ultrasound brain tissue incorporating expert evaluation. *Conf Proc IEEE Eng Med Biol Soc* 2005;6:6480–6483. [PubMed: 17281753]

**Table 1**

The percentage of scans in which each target finding was identified by the first, second and third readers. The third readers were sent only those films that the first or second reader thought abnormal. The last column lists the percentage of images in which each finding was agreed to by any two readers (concordant readings)

| Target finding                     | Reader |        |       | Final agreement |
|------------------------------------|--------|--------|-------|-----------------|
|                                    | First  | Second | Third |                 |
| Intraventricular hemorrhage        | 25     | 24     | 41    | 24              |
| Ventriculomegaly (moderate/severe) | 12     | 13     | 23    | 12              |
| Hyperechoic lesion                 | 23     | 22     | 27    | 16              |
| Hypoechoic lesion                  | 8.5    | 8.1    | 14    | 7.8             |

**Table 2**

The percentage of all scans read by pairs of readers that were read concordantly (positive/positive, negative/negative) and discordantly (positive/negative, negative/positive)

|               | Second reader                             |          | First reader |          |  |
|---------------|---|----------|--------------|----------|--|
|               | Positive                                  | Negative | Positive     | Negative |  |
|               | <b>Intraventricular hemorrhage</b>        |          |              |          |  |
| Positive      | 19  | 6        | 9            | 5        |  |
| Negative      | 6   | 69       | 4            | 83       |  |
| Agreement (%) | 76  | 92       | 69           | 94       |  |
| Kappa         | 0.68                                      |          | 0.63         |          |  |
|               | <b>Ventriculomegaly (moderate/severe)</b> |          |              |          |  |
| Positive      |   |          | 11           | 11       |  |
| Negative      |   |          | 13           | 65       |  |
| Agreement (%) |   |          | 46           | 86       |  |
| Kappa         |   |          | 0.32         |          |  |
|               | <b>Hyperechoic lesion</b>                 |          |              |          |  |
| Positive      |   |          | 5            | 5        |  |
| Negative      |   |          | 3            | 89       |  |
| Agreement (%) |   |          | 63           | 97       |  |
| Kappa         |   |          | 0.62         |          |  |

**Table 3**

Comparison between each reader and all other readers who read the same scans for the identification the four target findings. The *Less often* columns show the percentage of scan sets in which this reader did not identify the lesion when another reader did, and the *More often* columns show the percentage of scan sets in which this reader identified the lesion when the other reader did not

| Reader | Intraventricular hemorrhage |       |            | Moderate/severe ventriculomegaly |       |            | Hyperchoic lesion |       |            | Hypochoic lesion |       |            |
|--------|-----------------------------|-------|------------|----------------------------------|-------|------------|-------------------|-------|------------|------------------|-------|------------|
|        | Less often                  | Agree | More often | Less often                       | Agree | More often | Less often        | Agree | More often | Less often       | Agree | More often |
| A      | 6                           | 92    | 1          | 5                                | 92    | 2          | 18                | 82    | 0.5        | 1                | 98    | 1          |
| B      | 8                           | 88    | 4          | 8                                | 92    | 0.5        | 16                | 83    | 1          | 2                | 94    | 4          |
| C      | 14                          | 86    | 0.5        | 8                                | 89    | 3          | 19                | 71    | 10         | 3                | 94    | 3          |
| D      | 6                           | 88    | 5          | 5                                | 95    | 0.5        | 13                | 83    | 4          | 3                | 97    | 0          |
| E      | 6                           | 90    | 4          | 0.5                              | 96    | 4          | 10                | 86    | 4          | 3                | 94    | 3          |
| F      | 8                           | 87    | 5          | 4                                | 94    | 2          | 8                 | 68    | 24         | 4                | 94    | 2          |
| G      | 3                           | 89    | 8          | 1                                | 78    | 21         | 11                | 82    | 7          | 1                | 94    | 4          |
| H      | 2                           | 70    | 30         | 7                                | 88    | 5          | 10                | 70    | 20         | 0                | 95    | 5          |
| I      | 3                           | 86    | 11         | 5                                | 93    | 2          | 2                 | 52    | 45         | 5                | 94    | 1          |
| J      | 5                           | 93    | 2          | 1                                | 89    | 10         | 13                | 82    | 6          | 3                | 94    | 3          |
| K      | 2                           | 91    | 7          | 5                                | 90    | 4          | 13                | 86    | 1          | 3                | 94    | 2          |
| L      | 2                           | 86    | 12         | 1                                | 96    | 3          | 6                 | 75    | 19         | 2                | 94    | 3          |
| M      | 9                           | 89    | 2          | 4                                | 94    | 1          | 9                 | 68    | 23         | 4                | 90    | 7          |
| N      | 9                           | 90    | 1          | 3                                | 96    | 1          | 19                | 76    | 4          | 4                | 95    | 1          |
| O      | 5                           | 72    | 23         | 4                                | 94    | 2          | 10                | 69    | 22         | 4                | 93    | 4          |
| P      | 10                          | 83    | 7          | 12                               | 88    | 0          | 24                | 68    | 7          | 0                | 95    | 5          |
| Median | 5.5                         | 88    | 5          | 4                                | 92.5  | 2          | 10.5              | 75.5  | 6.5        | 3                | 94    | 3          |

**Table 4**

Analysis with four categories of ventriculomegaly. The larger the ventriculomegaly the higher the association with the other three target findings

| Most severe ventriculomegaly finding | Number of scan sets | Percentage of scan sets     |                    |                   |
|--------------------------------------|---------------------|-----------------------------|--------------------|-------------------|
|                                      |                     | Intraventricular hemorrhage | Hyperechoic lesion | Hypoechoic lesion |
| Normal                               | 2,038               | 8                           | 8                  | 3                 |
| Mild                                 | 494                 | 24                          | 12                 | 7                 |
| Moderate                             | 250                 | 40                          | 26                 | 15                |
| Severe                               | 118                 | 56                          | 54                 | 35                |
| Total number                         | 2,900               | 446                         | 350                | 170               |

**Table 5**

Concordance rates between the second and third readers and the first and third readers

| Target finding                     | Second reader | First reader |          | Third reader (% of disagreements read as positive) |
|------------------------------------|---------------|--------------|----------|--|
|                                    |               | Positive     | Negative |  |
| Intraventricular hemorrhage        | Positive      | 273          | 82       | 44   |
|                                    | Negative      | 94           | 1001     | 45   |
| Ventriculomegaly (moderate/severe) | Positive      | 125          | 66       | 36   |
|                                    | Negative      | 55           | 1204     | 42   |
| Hyperechoic lesion                 | Positive      | 156          | 165      | 19   |
|                                    | Negative      | 183          | 946      | 26   |
| Hypoechoic lesion                  | Positive      | 78           | 40       | 40   |
|                                    | Negative      | 45           | 1287     | 42   |

**Table 6**

The percentage of each set of protocol scans read by the first and second readers concordantly and discordantly for each target finding

| Target finding                     | Second reader |          | First reader |          | Scan 1 (n=1,134) |          | Scan 2 (n=1,344) |          | Scan 3 (n=1,252) |          |
|------------------------------------|---------------|----------|--------------|----------|------------------|----------|------------------|----------|------------------|----------|
|                                    | Positive      | Negative | Positive     | Negative | Positive         | Negative | Positive         | Negative | Positive         | Negative |
| Intraventricular hemorrhage        | 14            | 5        | 16           | 4        | 14               | 5        | 16               | 4        | 6                | 5        |
|                                    |               |          |              |          |                  |          |                  |          |                  |          |
|                                    | 4             | 16       | 6            | 74       | 4                | 16       | 6                | 74       | 4                | 86       |
| Ventriculomegaly (moderate/severe) | 3             | 4        | 7            | 4        | 3                | 4        | 7                | 4        | 6                | 2        |
|                                    |               |          |              |          |                  |          |                  |          |                  |          |
|                                    | 3             | 91       | 3            | 86       | 3                | 91       | 3                | 86       | 2                | 90       |
| Hyperchoic lesion                  | 8             | 9        | 8            | 8        | 8                | 9        | 8                | 8        | 2                | 6        |
|                                    |               |          |              |          |                  |          |                  |          |                  |          |
|                                    | 9             | 74       | 9            | 75       | 9                | 74       | 9                | 75       | 6                | 86       |
| Hypochoic lesion                   | 1             | 2        | 2            | 2        | 1                | 2        | 2                | 2        | 5                | 2        |
|                                    |               |          |              |          |                  |          |                  |          |                  |          |
|                                    | 1             | 96       | 1            | 95       | 1                | 96       | 1                | 95       | 2                | 90       |