



Published in final edited form as:

*Pediatr Res.* 2015 September ; 78(3): 304–309. doi:10.1038/pr.2015.99.

## Scoring system for periventricular leukomalacia in infants with congenital heart disease

Ann L. McCarthy<sup>1</sup>, Madeline E. Winters<sup>1</sup>, David R. Busch<sup>1,2</sup>, Ernesto Gonzalez-Giraldo<sup>3</sup>, Tiffany S. Ko<sup>4</sup>, Jennifer M. Lynch<sup>2</sup>, Peter J. Schwab<sup>1</sup>, Rui Xiao<sup>5</sup>, Erin M. Buckley<sup>6</sup>, Arastoo Vossough<sup>7</sup>, and Daniel J. Licht<sup>1</sup>

<sup>1</sup>Division of Neurology, The Children' Hospital of Philadelphia, Philadelphia, Pennsylvania

<sup>2</sup>Departments of Physics and Astronomy, University of Pennsylvania, Philadelphia, Pennsylvania

<sup>3</sup>Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania

<sup>4</sup>Bioengineering, University of Pennsylvania, Philadelphia, Pennsylvania

<sup>5</sup>Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, Pennsylvania

<sup>6</sup>The Optics Division, Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, Charlestown, Massachusetts

<sup>7</sup>Department of Radiology, The Children' Hospital of Philadelphia, Philadelphia, Pennsylvania

### Abstract

**Background**—Currently two magnetic resonance imaging (MRI) methods have been used to assess periventricular leukomalacia (PVL) severity in infants with congenital heart disease: manual volumetric lesion segmentation and an observational categorical scale. Volumetric classification is labor intensive and the categorical scale is quick but unreliable. We propose the quartered point system (QPS) as a novel, intuitive, time-efficient metric with high interrater agreement.

**Methods**—QPS is an observational scale that asks the rater to score MRIs on the basis of lesion size, number, and distribution. Pre- and postoperative brain MRIs were obtained on term congenital heart disease infants. Three independent observers scored PVL severity using all three methods: volumetric segmentation, categorical scale, and QPS.

**Results**—Hundred and thirty-five MRIs were obtained from 72 infants; PVL was seen in 48 MRIs. Volumetric measurements among the three raters were highly concordant ( $\rho = 0.94–0.96$ ). Categorical scale severity scores were in poor agreement between observers ( $\kappa = 0.17$ ) and fair agreement with volumetrically determined severity ( $\kappa = 0.26$ , respectively). QPS scores were in very good agreement between observers ( $\kappa = 0.82$ ) and with volumetric severity ( $\kappa = 0.81$ ).

**Conclusion**—QPS minimizes training and sophisticated radiologic analysis and increases interrater reliability. QPS offers greater sensitivity to stratify PVL severity and has the potential to more accurately correlate with neurodevelopmental outcomes.

---

Correspondence: Ann L. McCarthy (annlynmccarthy@gmail.com).

Disclosure: The authors have no financial relationships or conflicts of interest to disclose.

Periventricular leukomalacia (PVL) is the most common form of brain injury in preterm infants and is also commonly found in term infants with congenital heart disease (1–3). In broader context of premature injury, white matter injury (WMI) ranges from focal cystic necrosis to punctate white matter injury in the periventricular region (4–6). PVL occurs when immature glial precursor cells (premyelinating oligodendrocytes) within white matter are exposed to depressed oxygen levels and/or reduced blood flow (7). Newborns with PVL are at increased risk for permanent motor impairment and chronic neurological disabilities, including cerebral palsy, behavior deficits, and learning delays (8–11). The ability to assess the severity of PVL radiographically and to determine the relationship between imaging severity and clinical outcome is critical for physicians to provide proper guidance in long-term care and effective patient management.

PVL occurs in over half of the survivors of infant heart surgery and is a topic of intense research (12–16). Different techniques are currently used to evaluate PVL severity, namely, a popular observational scale we will call the ‘categorical’ scale, and volumetric ratings. Although several categorical systems exist, most are mainly modifications of a system proposed by Miller *et al.* (8), which stratifies lesions based on the number and size of white matter abnormalities seen as T1 hyperintense lesions on magnetic resonance imaging (MRI) (3,11,13). The categorical scale can be performed quickly and does not require sophisticated equipment for postprocessing images, but the severity scaling of PVL may be inaccurate. Alternatively, volumetric analysis has been used less frequently as it requires manual tracing of the T1 hyperintense lesions. This process is highly accurate and results in a continuous variable for analysis but is slow, very labor intensive, and requires a degree of training to perform. This volumetric method of quantifying PVL has become our group’s established standard measurement of WMI. With the limitations of both the volumetric and categorical approaches, a system that offers accurate scaling of volumetric analysis while preserving the simplicity of the categorical approach is needed. This study evaluates the two current methods used to measure PVL severity and we introduce a new, rapid, easy-to-use, and reproducible scoring system that can be readily available in both a clinical and research settings.

Of note, there is a general lack of consensus over the correct nomenclature for this type of injury. The T1 hyperintense lesions in the white matter have been labeled ‘punctate’ or ‘focal’ WMI depending largely on the research group. To be consistent with our previous publications, we consider WMI in the periventricular white matter watershed zone to be synonymous with PVL. Larger embolic infarcts, with cortical involvement and a well-defined vascular distribution, have a different injury profile on MRI and are excluded from this analysis. Currently, a consortium of researchers is working to develop standardized terminology to describe this WMI.

## RESULTS

A total of 72 patients were included in this study, resulting in 135 MRI scans (72 preoperative and 63 postoperative). All patients were full term with a mean gestational age at birth of  $38.9 \pm 1.0$  wk. PVL was seen in 13 (18%) of the preoperative scans and 35 (56%)

of the postoperative scans, yielding a total of 48 MRI scans which were used to evaluate all three scoring methods.

### Interrater Reliability Within Methods

**Volumetric measures**—The range of PVL lesion volumes was 0–9,309 mm<sup>3</sup>, and the medians for the three raters were 62, 78, and 63 mm<sup>3</sup>, respectively (Figure 1). All three observers' PVL volumetric measurements resulted in a concordance correlation coefficient range of  $\rho_c = 0.94$ –0.96 for pairwise comparison. There were seven scans with discordance among raters: six where two raters identified PVL and the third rater did not, and one where one rater identified PVL and the other two did not. Five of the seven scans had PVL lesion volumes of less than 31 mm<sup>3</sup>. Small lesions in regions of natural T1 hyperintensity (myelination) were particularly difficult to assess (Figure 2).

**Categorical scale**—The majority of the 48 PVL images were classified as “moderate” by all three raters, 90, 81, and 81%, respectively. However, one-third of the scans had discrepant scores. Interrater agreement between the three observers' scores resulted in a  $\kappa$  value of 0.17.

**Quartered point system**—The three raters classified most of the images similarly, with 39 (80%) scans receiving the identical score. The average quartered point system (QPS) scores for the three raters were 2.02, 2.08, and 1.98 (0–4-point scale). Details of the three raters' scores are shown in Figure 1. Interrater agreement between three observer's scores for the QPS resulted in a  $\kappa$  value of 0.82.

### Agreement Between Methods

**Categorical scale vs. volumetric measures**—The volumetric ground truth (averaged volume as assessed by three raters) ranged from 5.6 to 7,822 mm<sup>3</sup> with a median of 72 mm<sup>3</sup>. Twenty-four of the volumes were in the mild range (50%, 72mm<sup>3</sup>), 12 were in the moderate range (25%, >72–500 mm<sup>3</sup>), and 12 were severe (25%, >500 mm<sup>3</sup>).

The three raters' scores for the categorical scale were averaged together and resulted in the following scores: 4 mild (8%), 40 moderate (83%), and 4 severe (8%). Overall, the categorical scale was quite different from that of the volumetric ground truth volume scale; only 20 images (42%) were placed in the same categorical and volumetric group. The categorical scale was found to only be in fair agreement with the volumetric ground truth ( $\kappa = 0.26$ ; 95% CI: 0.12–0.39).

**Quartered point system vs. volumetric measures**—The three rater's QPS scores were averaged and resulted in the following scores: 1 none (2%), 20 mild (42%), 15 moderate (31%), and 12 severe (25%). Forty (83%) images were placed in the same QPS and volumetric group. For seven out of eight images where the QPS failed to assign the correct volumetric rating, the discrepancy was between the mild and moderate categories. The range of volumes for the seven discordant images was 39.4–99.1 mm<sup>3</sup> with an average volume of 65.8 mm<sup>3</sup>. The remaining discordant image was classified as none with QPS, as the three raters did not identify any PVL; however, a PVL volume of 5.6 mm<sup>3</sup> was identified

in the volume analysis (Figure 3). Ultimately, the volumetric ground truth and the QPS had similar distributions across all categories and were found to be in very good agreement ( $\kappa = 0.81$ ; 95% CI: 0.66–0.88).

The individual scores of each of the three raters were also in good or very good agreement with the volumetric ground truth (rater 1:  $\kappa = 0.83$ ; 95% CI: 0.78–0.84; rater 2:  $\kappa = 0.77$ ; 95% CI: 0.70–0.87; rater 3:  $\kappa = 0.79$ ; 95% CI: 0.68–0.89).

## DISCUSSION

Current literature has classified PVL severity without sophisticated radiologic tools; however, a reliable, standardized system has not been established. A new straightforward scoring system that accurately stratifies injury and correlates with the quantitative volumetric ground truth would be valuable in both the research and clinical settings. There are several qualities that a PVL severity scoring system needs to embody to reach broad acceptance. Key characteristics include a system that is quick and easy to use by a nonexpert, with results that can distribute across the scale and reflect quantitative findings.

We found that the QPS took less than 5 min to score each scan and was simple to use as it relies on number, size, and distribution of the lesions to assess the severity of PVL. The QPS guides the reviewer to ask four questions about the lesions, which when answered in stepwise fashion yield a score with excellent interrater reliability and accurately reflects the distribution of lesion volumes. There was some mild discordance around the breakpoint of mild and moderate (39.4–99.1 mm<sup>3</sup>; 65.8); however, the clinical significance of this overlap remains to be determined. The QPS eliminates the need for training or sophisticated radiologic analysis and provides a simple construct in which to stratify PVL severity in a highly accurate way.

The current categorical scales that are routinely used are quick and easy (less than 5 min per scan), yet do not offer the severity range that would afford higher-level analysis. Forty (83%) scans were classified as moderate by the categorical scale and spanned a very wide range of actual lesion volumes from 9.5 to 1,587.1 mm<sup>3</sup>. Figure 3 depicts the inordinate predominance of moderate severity scans with a wide range of volumes defined by the categorical scale. This predominance was largely due to the mild/moderate size cutoff point that placed any image with a lesion greater than 2 mm in the moderate category, which resulted in only four (8%) mild severity scans. In this cohort, most PVL lesions were 2 mm or larger making it difficult to stratify this population by severity, compromising the discriminatory value of this measurement. Using a 2 mm measurement effectively reduces PVL severity classification to a dichotomous variable of present versus absent and reduces the value of lesion dimensions as a surrogate marker. Thus, using the 5 mm threshold for lesion size is both easier for the rater to visualize (half a centimeter) and improves the discrimination between mild and moderate severities.

Additionally, the interrater reliability for the categorical scale was surprisingly low despite the predominance of scans classified as moderate. The low interrater reliability was mostly due to the misclassification of very small lesions and very severe lesions. As discussed

above, 2 mm is a relatively small measurement and one that is difficult to visualize (1/5th of a centimeter) while performing an observational scale. The scans with severe injury were also difficult to assess because of the subjective nature of determining percent of injury of the hemisphere (>5% of the hemisphere qualified as severe). As a consequence, none of the scans were scored as severe by more than one rater. Furthermore, when lesions were assessed volumetrically, no lesion, even the most severe, reached 5% of hemispheric volume.

The volumetric measurement is the most rigorous and yields continuous data; however, the process is tedious, labor intensive, and requires training. Acquisition time for the volumetric assessments was not measured but was estimated to range between 10 min for one scan with mild PVL to 25 min for a scan with more severe PVL. This does not include the time involved for training or software downloading and processing. Interrater reliability is good to very good and serves as our institution's ground truth measurement for PVL. However, this method may be difficult to disseminate for use in a multicenter study.

There are a number of limitations to this study. It is a single institutional experience and as such the imaging was standardized and performed on a single MRI scanner. Our findings will need to be validated by other groups. Neither volumetric or QPS measurements will be useful for rating cystic PVL given the anatomic distortion from the cysts. Although it has been shown that white matter abnormalities seen in term infants with congenital heart disease are identical (by pathology and radiographically) to those seen in premature infants, a similar study using MRIs obtained on a population of premature infants should be done to confirm the validity of the QPS (2,4,17). Another major limitation of this study is that it is a retrospective analysis that needs to be validated prospectively, optimally with neurodevelopmental outcomes. Clinical significance would first need to be established with the predetermined volumetric PVL cutoff points and subsequently with the QPS. A follow-up study with an expanded sample size (including premature infants) and raters from outside institutions could confirm our results and lead to the adoption of the QPS as a standardized tool to assess PVL severity.

## METHODS

### Neuroimaging

The current study is a re-analysis of MRI scans performed between 2008 and 2012, as part of a larger prospective study. The study was approved by the Institutional Review Board at The Children's Hospital of Philadelphia and parental informed consent was obtained. Patients with neonatal congenital heart disease requiring surgery in the first week of life were enrolled in this study. MRI scans were obtained immediately prior to surgery and again 7 to 10 d postoperatively. Inclusion and exclusion criteria are described in Licht *et al.* (18). Each scan was considered independently, raters did not have any knowledge of clinical data or whether the image was preoperative or postoperative.

MRI scans were performed using a Siemens (Erlangen, Germany) 1.5T Sonata scanner. Volumetric T1 MPRAGE (magnetization prepared rapid acquisition gradient echo) sequences in conjunction with diffusion-weighted imaging (DWI) were reviewed for

evidence of PVL. Voxel size of T1 MPRAGE's was  $0.41 \times 0.41 \times 1.5$  mm, and DWI voxel size was  $1.25 \times 1.25 \times 4$  mm. The presence of restricted diffusion on the DWI scans assisted in the confirmation of PVL when there was ambiguity in the T1 images. Note that DWI thickness and voxel size allows for higher spatial resolution resulting in greater correlation of T1 restricted diffusion and more accurate confirmation of PVL. The images were evaluated using three techniques to grade PVL severity: (i) volumetric measurement, (ii) a commonly used categorical scale, and (iii) a newly developed QPS. The skill levels of the raters were as follows: A.V. is an expert; A.L.M. and M.E.W. trained readers with a moderate level of expertise, and T.S.K. and E.G.-G. were novice readers. Not all raters reviewed all scans due, in part, to time constraints for some of the raters. Categorical and QPS analysis were done solely for this study, whereas volumetric was done routinely and in repetition by other raters throughout 2008–2012.

Volumetric measurements by manual lesion tracing and segmentation were performed by three independent observers (A.V., A.L.M., and M.E.W.). PVL lesions were manually segmented using ITK-SNAP version 3.2, which is an open-source imaging tool developed by Yushkevich (Pennsylvania, USA) and by Gerig (Utah, USA) and funded by the National Institute of Health (Maryland, USA) (19). The software has excellent intra- and interoperator reliability for measuring regional brain volumes (19) The software has excellent intra- and interoperator reliability for measuring regional brain volumes. Hyperintense lesions on T1 MPR images were identified, confirmed, and further segmented in three dimensions (Figure 4). Using the central sulcus (easily identified as T1 hyperintense relative to other cortical structures) to define anterior lesions from posterior lesions. PVL volumes were assessed in the whole brain as well as by dividing the brain into quadrants along the anterior–posterior axis and by hemisphere. PVL volumes are expressed in  $\text{mm}^3$ . MRI scans in which PVL was identified by at least one observer were included in the analysis. Ultimately, our volumetric standard (ground truth) was established as the average of the three observers' measured total volume and grouped, for statistical purposes, into predetermined volume categories.

PVL measurements greater than  $0 \text{ mm}^3$  and less than or equal to  $72 \text{ mm}^3$  (the median value for PVL volume, also used to define “clinically significant” PVL (20)) were considered mild, and those with volumes between  $72$  and  $500 \text{ mm}^3$  were categorized as moderate, and volumes  $>500 \text{ mm}^3$  were categorized as severe. The threshold for severe PVL was set at  $500 \text{ mm}^3$  because PVL volumes between  $237$ – $541 \text{ mm}^3$  were not represented in our cohort. These categories will require further validation against neurodevelopmental outcome variables.

Three independent observers (E.G.G., A.L.M., and M.E.W.) then evaluated the images with PVL using the commonly used categorical scale described by Miller *et al*: normal (no white matter lesions), minimal (three or fewer areas of T1 signal abnormality measuring  $<2$  mm), moderate ( $>3$  areas of T1 signal abnormality or these areas measuring  $>2$  mm, but  $<5\%$  of the hemisphere involved), or severe ( $>5\%$  of the hemisphere involved) (8) and assigned each a numerical value (0–3; Table 1).

Finally, three independent observers (T.S.K., A.L.M., and M.E.W.) assessed the images using the QPS tool. For the assessment, four questions were posed for each image: (i) Is PVL present? (ii) Is there PVL in all four quadrants? (iii) Are there three or more lesions in one quadrant? and (iv) Are any of the lesions greater than 5 mm? One point is given for each affirmative answer, resulting in a 0–4-point scale (Table 2). For the purposes of QPS, an observational rating system, lesions of 5 mm (half a centimeter), were easier to visualize than lesions of 2 mm (1/5th of a centimeter). Quadrants were defined following the same criteria used for the PVL volumetric assessment.

### Statistical Analysis

Interrater reliability was determined for each of the three PVL scoring methods. The concordance correlation coefficient was assessed for pairwise agreement among the three observers' continuous volumetric measurements, and the Fleiss'  $\kappa$  statistic was used for the interrater reliability of the categorical scale and QPS.

To assess the agreement between the three approaches within each method, the three observers' scores were averaged together for each patient's imaging data. This resulted in one averaged PVL volume (ground truth volume), one averaged categorical score, and one averaged QPS score for each image. The continuous volumetric data was then converted to categorical data using volume cutoffs as described in the Methods section (Table 1). Cohen's  $\kappa$  statistics were used to compare the volumetric ground truth to the categorical score and QPS. All  $\kappa$  statistics were linearly weighted to account for the degree of disagreement between the categories. Altman's qualitative system for strength of agreement for kappa statistics was used to describe interrater reliability (21). Data were analyzed with Stata software version 10 (StataCorp, College Station, TX), and statistical significance was assessed at the 0.05 level. Study data were collected and managed using REDCap electronic data capture tools hosted at The Children's Hospital of Philadelphia.

### Conclusions

The quartered point system is a simple and accurate method that clinicians and researchers can use to determine the severity of PVL. It is in very good agreement with the time-intensive ground truth of direct volume measurement and has the potential to provide physicians with the tools to make a more timely assessment. Due to the greater sensitivity to the range of severity of PVL, QPS has a greater potential to more accurately correlate with neurodevelopmental outcomes.

### Acknowledgments

#### STATEMENT OF FINANCIAL SUPPORT

This study was supported by: a 2013 Student Scholarship in Cerebrovascular Disease and Stroke at the American Heart Association's Stroke Council, Dallas, TX (A.L.M.); National Heart, Lung and Blood Institute HL007915, Bethesda, MD (D.R.B.); Thrasher Research Foundation Early Career Award, Salt Lake City, UT (D.R.B.); National Institutes of Health T32 grant no. HL007954, Bethesda, MD (J.M.L.); National Institutes of Health grant no. NS-072338, Bethesda, MD (D.J.L.); and the June and Steve Wolfson Family Foundation, Philadelphia, PA (D.J.L.).

## References

1. Inder T, Huppi PS, Zientara GP, et al. Early detection of periventricular leukomalacia by diffusion-weighted magnetic resonance imaging techniques. *J Pediatr*. 1999; 134:631–634. [PubMed: 10228300]
2. Miller SP, McQuillen PS, Hamrick S, et al. Abnormal brain development in newborns with congenital heart disease. *N Engl J Med*. 2007; 357:1928–1938. [PubMed: 17989385]
3. Galli KK, Zimmerman RA, Jarvik GP, et al. Periventricular leukomalacia is common after neonatal cardiac surgery. *J Thorac Cardiovasc Surg*. 2004; 127:692–704. [PubMed: 15001897]
4. Kinney HC, Back SA. Human oligodendroglial development: relationship to periventricular leukomalacia. *Semin Pediatr Neurol*. 1998; 5:180–189. [PubMed: 9777676]
5. Volpe JJ. Neurobiology of periventricular leukomalacia in the premature infant. *Pediatr Res*. 2001; 50:553–562. [PubMed: 11641446]
6. Back SA, Rivkees SA. Emerging concepts in periventricular white matter injury. *Semin Perinatol*. 2004; 28:405–414. [PubMed: 15693397]
7. Back SA, Han BH, Luo NL, et al. Selective vulnerability of late oligodendrocyte progenitors to hypoxia-ischemia. *J Neurosci*. 2002; 22:455–463. [PubMed: 11784790]
8. Miller SP, Ferriero DM, Leonard C, et al. Early brain injury in premature newborns detected with magnetic resonance imaging is associated with adverse early neurodevelopmental outcome. *J Pediatr*. 2005; 147:609–616. [PubMed: 16291350]
9. Back SA. Perinatal white matter injury: the changing spectrum of pathology and emerging insights into pathogenetic mechanisms. *Ment Retard Dev Disabil Res Rev*. 2006; 12:129–140. [PubMed: 16807910]
10. Jalali A, Buckley EM, Lynch JM, Schwab PJ, Licht DJ, Nataraj C. Prediction of periventricular leukomalacia occurrence in neonates after heart surgery. *IEEE J Biomed Health Inform*. 2014; 18:1453–1460. [PubMed: 24122606]
11. Rios DR, Welty SE, Gunn JK, et al. Usefulness of routine head ultrasound scans before surgery for congenital heart disease. *Pediatrics*. 2013; 131:e1765–e1770.
12. Beca J, Gunn JK, Coleman L, et al. New white matter brain injury after infant heart surgery is associated with diagnostic group and the use of circulatory arrest. *Circulation*. 2013; 127:971–979. [PubMed: 23371931]
13. Andropoulos DB, Hunter JV, Nelson DP, et al. Brain immaturity is associated with brain injury before and after neonatal cardiac surgery with high-flow bypass and cerebral oxygenation monitoring. *J Thorac Cardiovasc Surg*. 2010; 139:543–556. [PubMed: 19909994]
14. Limperopoulos C, Majnemer A, Shevell MI, et al. Predictors of developmental disabilities after open heart surgery in young children with congenital heart defects. *J Pediatr*. 2001; 141:51–58. [PubMed: 12091851]
15. Goff DA, Shera DM, Tang S, et al. Risk factors for preoperative periventricular leukomalacia in term neonates with hypoplastic left heart syndrome are patient related. *J Thorac Cardiovasc Surg*. 2014; 147:1312–1318. [PubMed: 23879933]
16. Mahle WT, Tavani F, Zimmerman RA, et al. An MRI study of neurological injury before and after congenital heart surgery. *Circulation*. 2002; 106(Suppl 1):I109–I114. [PubMed: 12354718]
17. Glauser TA, Rorke LB, Weinberg PM, Clancy RR. Acquired neuropathologic lesions associated with the hypoplastic left heart syndrome. *Pediatrics*. 1990; 85:991–1000. [PubMed: 2339048]
18. Licht DJ, Shera DM, Clancy RR, et al. Brain maturation is delayed in infants with complex congenital heart defects. *J Thorac Cardiovasc Surg*. 2009; 137:529–536. discussion 536–7. [PubMed: 19258059]
19. Yushkevich PA, Piven J, Hazlett HC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage*. 2006; 31:1116–1128. [PubMed: 16545965]
20. Lynch JM, Buckley EM, Schwab PJ, et al. Time to surgery and preoperative cerebral hemodynamics predict postoperative white matter injury in neonates with hypoplastic left heart syndrome. *J Thorac Cardiovasc Surg*. 2014; 148:2181–2188. [PubMed: 25109755]

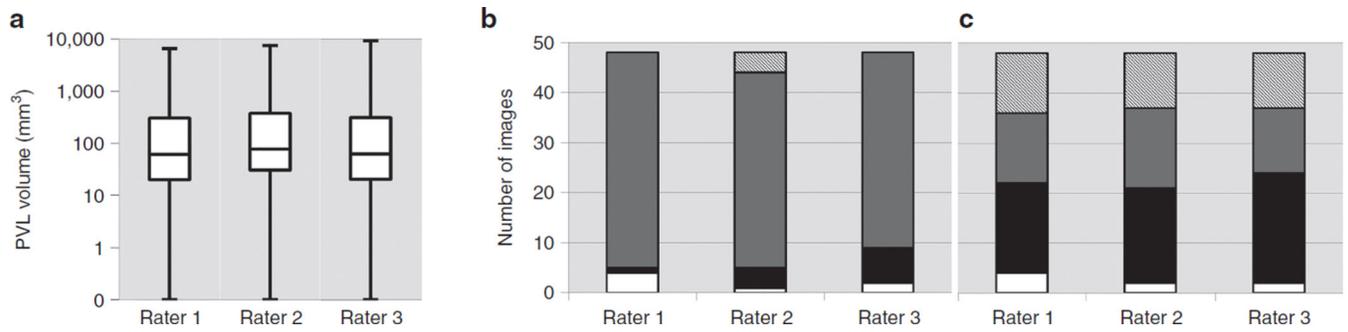
21. Altman, DG. Practical Statistics for Medical Research. 1st. Boca Raton, FL: Chapman and Hall/CRC Press; 1990. p. 403-409.

Author Manuscript

Author Manuscript

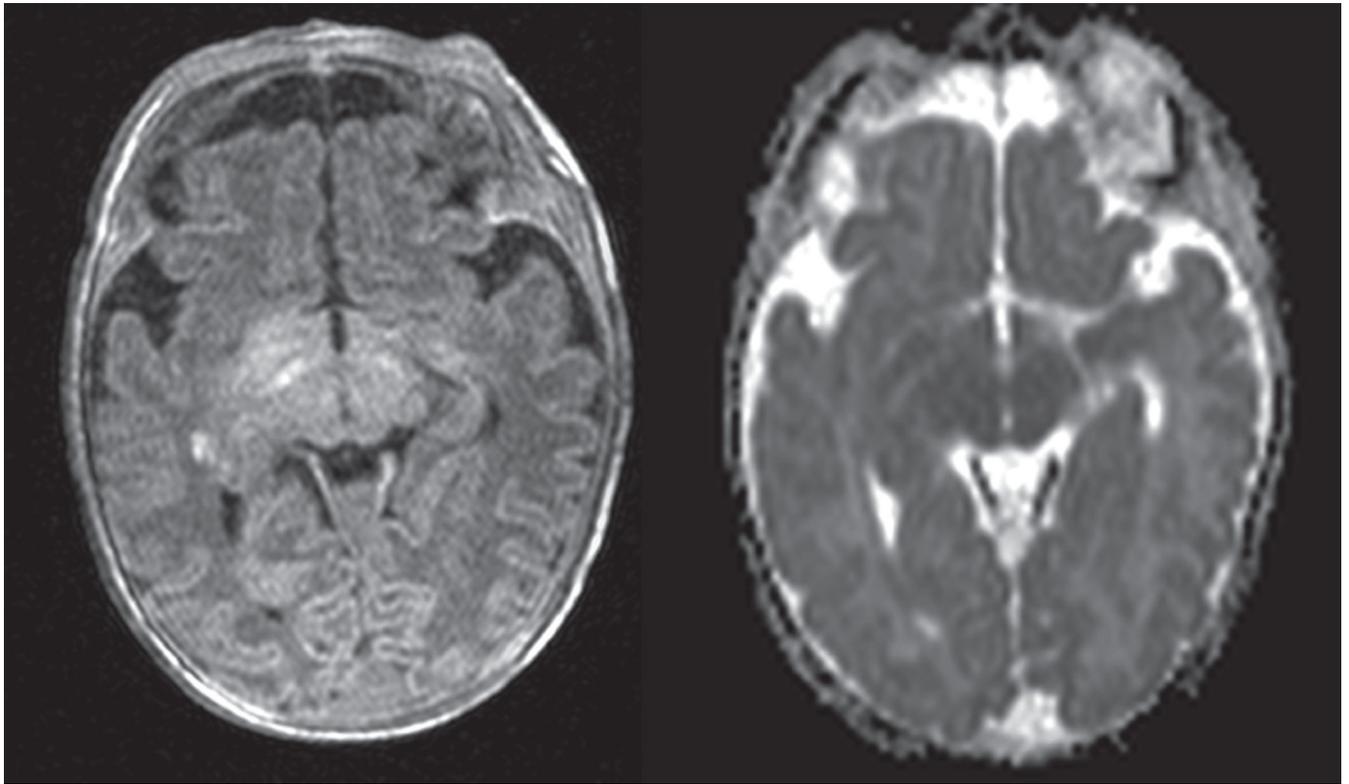
Author Manuscript

Author Manuscript

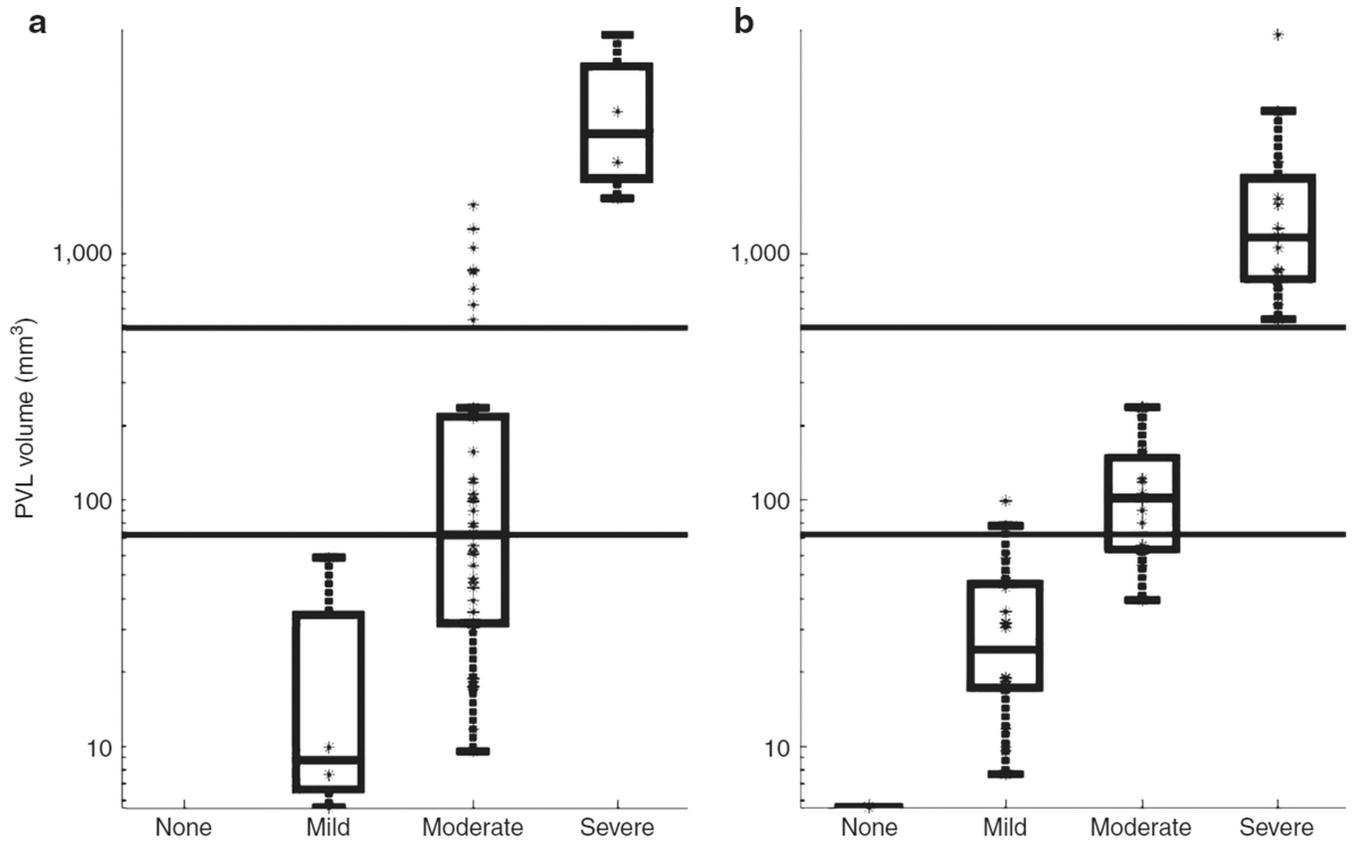


**Figure 1.**

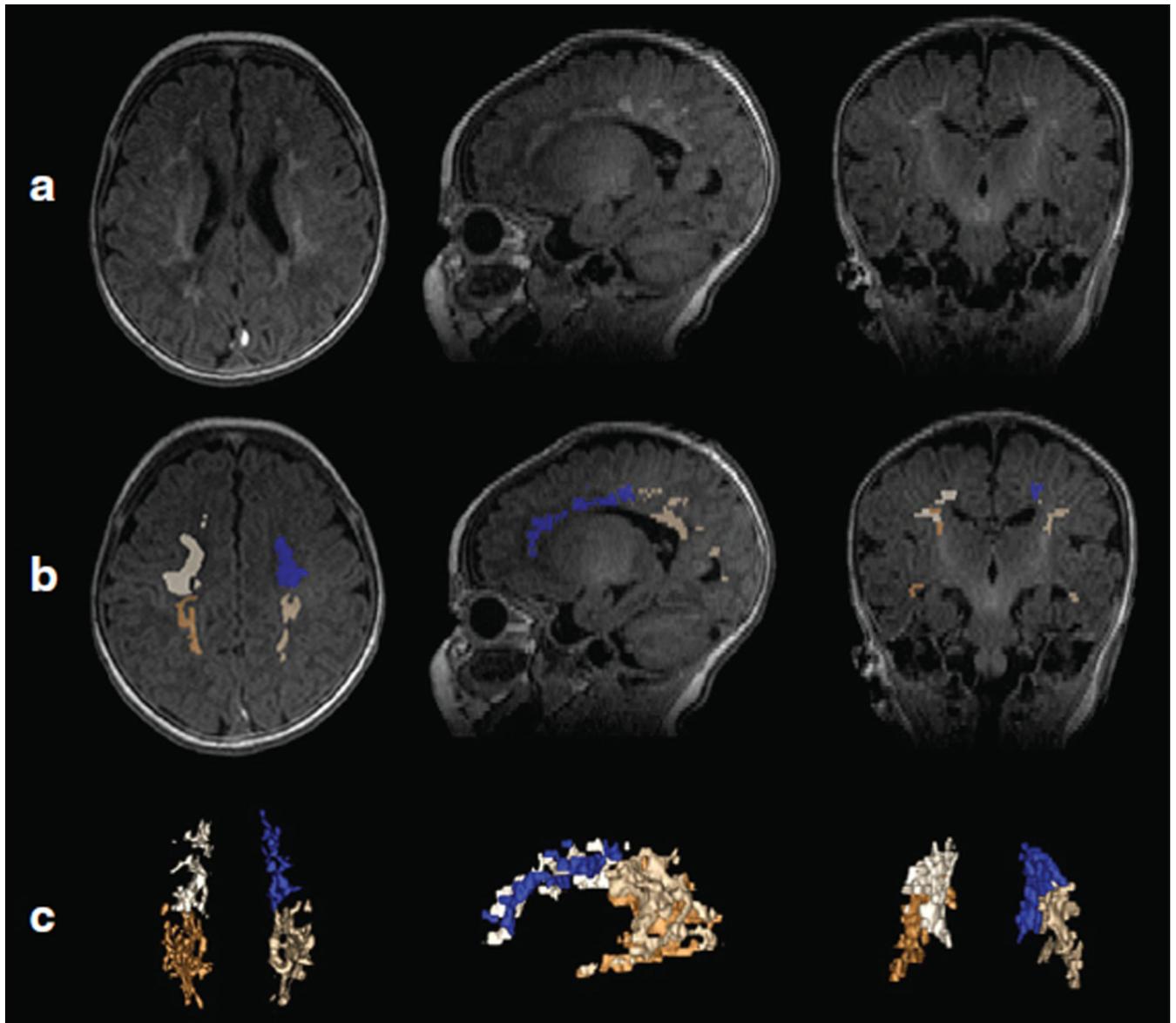
Rater results. Results from each rater are shown for all three scoring methods: (a) volumetric measurement (log scale), (b) categorical scale, and (c) quartered point system (QPS). In a, the range of PVL volumes was 0–9,309 mm<sup>3</sup> and the medians for the three raters were 62, 78, and 63 mm<sup>3</sup>, respectively. Whiskers denote the respective minimum and maximum volumes measured. In b and c, the distribution of PVL severity graded “none” (white), “mild” (black), “moderate” (gray), and “severe” (hatched) for each rater is shown using the (b) categorical scale and (c) QPS. Using the categorical scale, the majority of the 48 images were classified as “moderate” by all three raters (i.e., 90, 81, and 81%, respectively). While a third or less of images were classified as “moderate” by raters using QPS (29, 33, and 29% respectively). PVL, periventricular leukomalacia.



**Figure 2.** Rater discrepancy. T1 (left) and DWI (right) scans where motion degradation of the image resulted in discrepancies in the interpretation of PVL volumes. DWI, diffusion-weighted imaging; PVL, periventricular leukomalacia.



**Figure 3.** Comparison of three scoring systems. Mean volumetric measurements for each scan are represented logarithmically along the y-axis. Bold horizontal lines indicate the thresholds used to determine PVL severity volumetrically: “mild” ( 72 mm<sup>3</sup>), “moderate” ( 500 mm<sup>3</sup>), and “severe” (>500 mm<sup>3</sup>). The volume distribution of the (a) categorical scale shows notably higher aggregation of volumetric severity within grades. In contrast, the (b) quartered point system exhibits an improved representation of volumetric severity. Forty (83%) scans were classified as moderate by the categorical scale and spanned a very wide range of actual lesion volumes. PVL, periventricular leukomalacia.



**Figure 4.** Three dimensional segmentation. (a) The T1-weighted image prior to manual PVL segmentation (b) the manual segmentation, and (c) the volumetric reconstruction colored by quadrant, are shown in the transverse (left), sagittal (middle), and coronal (right) plane for a subject graded “severe” by all three scoring systems. PVL, periventricular leukomalacia.

**Table 1**

## PVL scoring systems

Category	Volumetric measurement (mm <sup>3</sup> )	Categorical scale	Quartered point system
None	0	No lesions observed	0
Mild	>0–72	3 areas of abnormality, each smaller than 2 mm	1
Moderate	>72–500	>3 areas of abnormality or areas larger than 2 mm	2–3
Severe	>500	>5% of the involved hemisphere	4

PVL, periventricular leukomalacia.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2**

Quartered point system

---

**QPS questions (if yes, one point for each)**

---

Is PVL present?

Is there PVL in all four quadrants?

Are there three or more lesions in one quadrant?

---

Are any of the lesions greater than 5 mm?

---

PVL, periventricular leukomalacia.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript