

ORIGINAL ARTICLE

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?

D L Harris, R L Teele, F H Bloomfield, J E Harding, on behalf of the Australian and New Zealand Neonatal Network

Arch Dis Child Fetal Neonatal Ed 2005;90:F494–F499. doi: 10.1136/adc.2004.065219

Background: The incidence of germinal matrix/intraventricular haemorrhage (GM/IVH) reported to the Australian and New Zealand Neonatal Network (ANZNN) varies between neonatal intensive care units (NICUs).

Hypothesis: Differences in the capture, storage, and interpretation of the cerebral ultrasound scans may account for some of this variation.

Methods: A total of 255 infants with birth weight <1500 g and gestation <32 weeks born between 1997 and 2002 were randomly selected from the ANZNN database, 44 from each of the six NICUs in New Zealand. Twenty two infants from each NICU had cerebral ultrasound scans previously reported to ANZNN as normal; another 22 had scans reported as abnormal. The original scans were copied using digital photography and anonymised and independently read by a panel of three experts using a standardised method of reviewing and reporting.

Results: There was considerable variation between NICUs in methods of image capture and quality and completeness of the scans. However, there was little variation in the reporting of scans between the reviewers and the reports to ANZNN (weighted κ 0.75–0.91). Grade 1 GM/IVH was generally over-reported and grade 4 under-reported to the ANZNN.

Conclusion: For all NICUs, a high level of agreement was found between the reviewers' reports and the reports to the ANZNN. Thus the variation between NICUs in the incidence of GM/IVH reported to the ANZNN is unlikely to be due to differences in capture, storage, and interpretation of the cerebral ultrasound scans. Further investigation is warranted into the reasons for the variation in incidence of GM/IVH between NICUs.

See end of article for authors' affiliations

Correspondence to:
Professor Harding, Liggins Institute, Faculty of Medicine and Health Science, University of Auckland, Private Bag 92019, Auckland, New Zealand; j.harding@auckland.ac.nz

Accepted 18 March 2005

Germinal matrix/intraventricular haemorrhage (GM/IVH) is the most common type of intracranial haemorrhage in the neonatal period, and contributes to long term neurological problems such as cerebral palsy and hydrocephalus.¹ The incidence and severity of GM/IVH is closely related to gestational age and birth weight, occurring in 25–30% of babies born at less than 1500 g or less than 32 weeks gestation.

International neonatal databases have consistently reported wide variation in the incidence of clinical outcomes, including GM/IVH, between different neonatal intensive care units (NICUs).^{2–5} In 1993 the International Neonatal Network called for comparisons of staffing and organisational policies of NICUs to investigate reasons for these variations.⁶

The Australian and New Zealand Neonatal Network (ANZNN) collates data from all 29 level 3 NICUs in Australia and New Zealand. Since 1995 each NICU has collected a dataset of 60 variables, using agreed definitions, on all babies born before 32 weeks gestation or with a birth weight less than 1500 g, and all babies needing major surgery or requiring assisted ventilation for over four hours. Grading of GM/IVH follows the scale established by Papile, and the highest diagnosed grade of GM/IVH is reported.⁷

In the 1995 dataset, the reported rates of severe GM/IVH varied from less than 5% to 20% in the different NICUs.⁸ This reported variation may be influenced by a number of factors.⁵ There is often wide variation between NICUs in case mix,

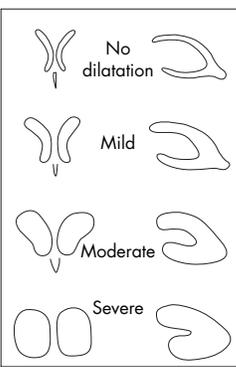
infant risk factors, and the number of neonates, all of which may have an impact on both mortality and clinical outcome.⁹ In addition, measurement bias in assessing the outcome, such as differences in techniques and interpretation of diagnostic tests, may affect the reported variation.

To allow appropriate adjustment for risk factors in different NICUs, a recent study established a predictive model for major GM/IVH based on the ANZNN database.⁸ However, when the predictive risk model was used to adjust for risk factors within each NICU population, the variation between NICUs in the reported incidence of GM/IVH increased rather than decreased.¹⁰

An alternative explanation for the differences in reported incidence of GM/IVH is that the scans are performed, documented, and reported differently in each NICU. Our study was undertaken to test the hypothesis that the reported variation in the incidence of GM/IVH was due to differences in the capture and interpretation of cerebral ultrasound scans. As part of this study, we also assessed both intraobserver and interobserver variability in the reporting of cerebral ultrasound scans.

Abbreviations: ANZNN, Australian and New Zealand Neonatal Network; GM/IVH, germinal matrix/intraventricular haemorrhage; NICU, neonatal intensive care unit

The Peri/Intraventricular Haemorrhage verification study

| | | |
|--|--|---|
| <p>Machine used to scan: Early: Acuson / ATL / GE / Toshiba / Other _____ Late: Acuson / ATL / GE / Toshiba / Other _____</p> <p>Probe used to scan: Early: _____ Late: _____</p> <p>View of images: Early: anterior / posterior Late: anterior / posterior fontanelle</p> <p>Completeness of scan: Early: incomplete / complete ie 6 coronal, 5 sagittal views Late: incomplete / complete ie 6 coronal, 5 sagittal views</p> <p>Quality of scan: Early: good / poor (compromises interpretation of scan) Late: good / poor (compromises interpretation of scan)</p> <p>Comments: _____</p> |  | <p>Study number: _____</p> <p>NZ assessor: FB / DH / RT</p> <p>Date of assessment: ___ / ___ / ___</p> <p>Gestational age: 23 24 25 26 27 28 29 30 31 weeks</p> |
|--|--|---|

| Patient age (days) | Normal scan | Blood in sub-ependyma | Ventricles | | | | | Parenchyma | | | Comments |
|--------------------|-------------|-----------------------|--------------------|-----------|----------------|------------------------------|------------------------|------------|----------------------|-------|----------|
| | | | Blood in ventricle | Distended | Distended with | Lateral ventricle size at FM | If dilatation | Blood | White matter changes | Cysts | |
| Early scan | Yes | Yes | Yes | Yes | Blood CSF | Normal | Unilat. R | Yes | None | Yes | |
| | No | No | No | No | Both | Mild Moderate Severe | Unilat. L Bilateral | No | Single Multiple | No | |
| Late scan | Yes | Yes | Yes | Yes | Blood CSF | Normal | Unilat. R | Yes | None | Yes | |
| | No | No | No | No | Both | Mild Moderate Severe | Unilat. L Bilateral | No | Single Multiple | No | |

Figure 1 Data collection sheet used for reporting of cerebral ultrasound scans for this study.

METHODS

The study population was randomly selected from the New Zealand subset of the ANZNN database for 1997–2002. Only babies from the six level 3 NICUs in New Zealand were included to allow population based data and easier access to all of the original scans. All babies met the high risk criteria for GM/IVH in that they were born at <32 weeks gestation or weighing <1500 g. Forty four babies from each NICU were selected by ANZNN staff using a computerised three digit, random number table. Half of each group had cerebral ultrasound scans that were reported as normal, one quarter were reported as having grade 1 or 2 GM/IVH, and the final quarter grade 3 or 4 GM/IVH. The reviewers were unaware of this selection pattern. When an ultrasound scan was unavailable because, for example, the films were lost or had been culled from hospital files, the baby was replaced from the ANZNN database with another baby from the same NICU with the same reported cerebral ultrasound result whenever possible.

Depending on the hospital, the cerebral ultrasound scans were stored as digital images, on thermal paper, or on x ray film. The hard copy cerebral ultrasounds were photographed using a digital camera and a standardised protocol and stored digitally as Joint Photographic Experts Group (JPEG) files. All identifying information was deleted from digital images using PHOTOSHOP 5.0 LE (Adobe Systems Inc, San Jose, California, USA) by an assistant who was not involved in any other aspect of the study. The resulting images were stored as Photoshop document files.

The definition used for reporting GM/IVH to the ANZNN is the worst level of GM/IVH seen on either side of the brain by ultrasonography or post mortem examination.¹¹ As it was impossible to know which particular cerebral ultrasound scan was originally considered the worst and reported by the NICU to the ANZNN database, all scans performed between 4 and 14 days of age were copied for each baby (unless the baby had died, in which case all available scans were copied).

The three reviewers were a specialist paediatric radiologist, a neonatologist, and a neonatal nurse practitioner, who are all required to perform and report cerebral ultrasound scans as part of their clinical practice. The reviewers were unaware of the original reports. The only clinical information provided was the gestational age of each baby. The reviewers were asked to evaluate all photographed scans in a dimmed room with a standard computer screen and to report on only the worst scan for each baby. The reviewers were not asked to report a grade of GM/IVH. Rather, each reviewer completed a data collection sheet (fig 1) noting the presence of blood within the germinal matrix or ventricles, ventricular distension, and parenchymal abnormality. Parenchymal abnormality was classified as blood, white matter changes, or cysts. These data were then converted into Papile’s grading method using a computer algorithm (table 1).⁷

A complete scan was defined as a scan that included the standard six coronal views and five sagittal views.^[12 13] If there were any views missing, then the scan was defined as incomplete. The quality of the scan was defined as good if it allowed interpretation. Otherwise the scan was defined as

Table 1 Algorithm used to convert the reviewers' reports into Papile's grading system

| Recorded on the data collection sheet | Normal | Grade 1 | Grade 2 | Grade 3 | Grade 4 |
|---------------------------------------|--------|---------|---------|---------|---------|
| Normal scan | Yes | No | No | No | No |
| Subependymal haemorrhage | No | Yes | Yes | Yes | Yes |
| Blood in ventricle | No | No | Yes | Yes | |
| Ventricle distended with blood | No | No | No | Yes | |
| Parenchymal abnormality | No | No | No | No | Yes |

poor. Scans that were unable to be categorised by the Papile grading method using a computer algorithm were tabulated as not graded. Scans that were impossible for the reviewers to read were also tabulated.

To assess intraobserver variability, all three reviewers reviewed again a randomly selected subset of cerebral ultrasound scans from 22 babies two months after the initial review.

The chairperson of the Auckland ethics committee confirmed in writing that there was no need for ethical approval for this study as it constitutes a clinical audit. All identifying data were removed from the copied films, and there was no contact with any of the infants or their families.

All data were analysed using Statistical Discovery Software Jmp 5. (SAS Institute Inc, Cary, North Carolina, USA). Gestation and birth weight were compared between NICUs using factorial analysis of variance. The incidences of different abnormalities were compared between reviewers and NICUs using χ^2 . Interobserver and intraobserver variability were assessed using the weighted kappa statistic, $\kappa (w)$. All data are presented as number (%) or median (range).

RESULTS

A total of 2225 cerebral ultrasound scans from 255 babies were copied, then reviewed by the panel of three experts. The median gestational age of the babies was 28 weeks (range 22–31), and the median birth weight was 1025 g (470–2140). In this sample the incidence of normal scans reported to the ANZNN from different NICUs ranged from 45% to 55%, and of grade 4 GM/IVH from 8% to 23%.

Ultrasound equipment, probe sizes, and scanning techniques varied among NICUs (table 2). Sixty nine percent of the scans were taken via the anterior fontanel, 28% via both the anterior and posterior fontanels, and 3% were axial views.

Quality and completeness

Thirty seven percent of scans were considered incomplete, and 33% were of poor quality. There was significant variation between the NICUs in the quality and completeness of the cerebral ultrasound scans ($p < 0.001$). NICUs E and F had 90% of the incomplete and 70% of the poor quality scans, and it was only scans from these two NICUs that the reviewers found impossible to read (table 3).

Variation between NICUs

There was some variation in the reporting of GM/IVH between the reviewers and the ANZNN for each NICU. If the reviewers are considered the gold standard, then grade 1 GM/IVHs were over-reported to the ANZNN by 5–12% by every unit apart from NICU F. Grade 2 GM/IVH had less variation between the reviewers and the ANZNN reports. Grade 3 GM/IVH showed the greatest variation in reporting, being both under-reported by 11% and over-reported by 10%. Grade 4 was under-reported by every NICU by 1–11% (table 4).

Intraobserver variability

When the original reports for each reviewer were compared with the reports from the re-read subset, we found that the reports did not change, with the $\kappa (w)$ statistics ranging from 0.78 to 0.96 (table 5).

Interobserver variability

The reviewers did not differ in their overall reporting of numbers of scans that were incomplete, poor quality, or impossible to read (table 6). However, they did not always agree on which specific scans were in each of these categories. For this reason the reports from each reviewer

Table 2 Ultrasound equipment and technique used in each neonatal intensive care unit (A–F)

| | A | B | C | D | E | F |
|--------------------|------------------------------------|---------------------|---------------------|---------------------|---------------------|----------------------------|
| Ultrasound machine | Acuson Aspen, Siemens, Sequoia 512 | Acuson Aspen | ATL HDI 3000 | ATL HDI 3000 | ATL HDI 3000 | Acuson Aspen, ATL HDI 3000 |
| Probe size (MHz) | 7–8.5 | 5–7 | 8.5 | 7–8.5 | 7.5–8.5 | 5–7.5 |
| Views | Anterior, posterior | Anterior, posterior | Anterior, posterior | Anterior, posterior | Anterior, posterior | Anterior, posterior, axial |

Not all views were used in all cases.

Table 3 Variation between neonatal intensive care units (NICUs) A–F in the reported quality and completeness of scans

| | A | B | C | D | E | F | Total |
|--------------------|---------|---------|---------|-------|----------|----------|----------|
| Total babies | 42 | 42 | 44 | 40 | 43 | 44 | 255 |
| Total reports | 126 | 126 | 132 | 120 | 129 | 132 | 765 |
| Incomplete scans | 15 (12) | 13 (11) | 4 (3) | 7 (6) | 128 (99) | 107 (85) | 274 (37) |
| Poor scans | 10 (8) | 25 (20) | 35 (27) | 5 (4) | 114 (88) | 59 (47) | 248 (33) |
| Impossible to read | 0 | 0 | 0 | 0 | 36 (28) | 4 (3) | 40 (5) |
| Not graded | 15 (12) | 3 (2) | 1 (1) | 3 (2) | 6 (5) | 6 (5) | 34 (4) |
| Missing scans | 0 | 6 (5) | 0 | 0 | 0 | 6 (5) | 12 (2) |

The proportion of scans that were incomplete or of poor quality varied between NICUs, $p < 0.001$. The values are number (%) of reports for each NICU. Scans may be included in more than one category—for example, because the reviewer was able to report on the appearance of the ventricles but not the parenchyma.

Table 4 Comparison between the three reviewers and the ANZNN in the reported grades of germinal matrix/intraventricular haemorrhage for each neonatal intensive care unit (A–F)

| | A | B | C | D | E | F | Total |
|-----------|---------|---------|---------|---------|---------|---------|----------|
| Babies | 42 | 42 | 44 | 40 | 43 | 44 | 255 |
| Reports | | | | | | | |
| Reviewers | 111 | 117 | 131 | 117 | 87 | 116 | 679 |
| ANZNN | 37 | 39 | 44 | 40 | 33 | 39 | 232 |
| Normal | | | | | | | |
| Reviewers | 66 (59) | 78 (67) | 77 (60) | 59 (50) | 39 (45) | 55 (47) | 374 (56) |
| ANZNN | 19 (51) | 21 (54) | 24 (55) | 20 (51) | 15 (45) | 20 (51) | 119 (52) |
| Grade 1 | | | | | | | |
| Reviewers | 9 (8) | 3 (3) | 6 (5) | 8 (7) | 6 (7) | 13 (11) | 45 (7) |
| ANZNN | 5 (14) | 6 (15) | 7 (16) | 7 (15) | 4 (12) | 3 (8) | 31 (13) |
| Grade 2 | | | | | | | |
| Reviewers | 12 (11) | 11 (9) | 13 (9) | 12 (11) | 4 (5) | 9 (8) | 61 (8) |
| ANZNN | 4 (10) | 3 (9) | 3 (7) | 3 (8) | 4 (12) | 6 (15) | 23 (10) |
| Grade 3 | | | | | | | |
| Reviewers | 7 (6) | 6 (5) | 11 (9) | 8 (6) | 17 (20) | 15 (13) | 64 (9) |
| ANZNN | 6 (16) | 3 (8) | 6 (14) | 1 (2) | 3 (9) | 6 (15) | 24 (10) |
| Grade 4 | | | | | | | |
| Reviewers | 17 (15) | 19 (16) | 24 (17) | 30 (26) | 21 (25) | 24 (21) | 133 (20) |
| ANZNN | 3 (8) | 6 (15) | 4 (9) | 9 (23) | 7 (21) | 4 (10) | 33 (14) |

Values are total number (%) of reports. Scans (n = 86) that were reported as impossible to read, not graded, or missing have been excluded.

Table 5 Variation in the reported grades of germinal matrix/intraventricular haemorrhage in the re-read subset of scans (indicated by the R column for each reviewer X–Z)

| | X | XR | Y | YR | Z | ZR |
|---------------|---------|--------|---------|--------|---------|--------|
| Total babies | 22 | 22 | 22 | 22 | 22 | 22 |
| Total reports | 20 | 15 | 19 | 13 | 16 | 16 |
| Normal | 12 (60) | 9 (60) | 12 (63) | 7 (54) | 10 (63) | 9 (56) |
| Grade 1 | 2 (10) | 0 | 0 | 0 | 1 (6) | 2 (13) |
| Grade 2 | 1 (5) | 2 (13) | 2 (11) | 0 | 2 (13) | 1 (6) |
| Grade 3 | 1 (5) | 1 (7) | 1 (5) | 1 (8) | 0 | 0 |
| Grade 4 | 4 (20) | 3 (20) | 4 (21) | 5 (38) | 3 (19) | 4 (25) |

Values are numbers (%) of reports. Scans (n = 33) that were reported as impossible to read, not graded, or missing have been excluded.

were compared individually with those reported to ANZNN. There were no consistent differences between reviewers in the reporting of grades of GM/IVH (table 7). κ (w) statistics ranged from 0.75 to 0.91, and all 95% confidence intervals overlapped, showing no significant differences in the level of agreement between any sources of reports (table 8).

DISCUSSION

We sought to discover whether some of the variation in the incidence of GM/IVH reported from the six NICUs in New Zealand could be accounted for by differences in the cerebral ultrasound scanning method and interpretation. We found that there was significant variation in the quality and completeness of cerebral ultrasound scans from each of the NICUs in New Zealand. Despite this, there was a very high

level of agreement between the reviewers’ reports and the NICU reports to the ANZNN, and very low levels of intraobserver and interobserver variability. Thus variations in cerebral ultrasound scanning method and interpretation were unlikely to explain the reported variation in the incidence of GM/IVH between New Zealand NICUs.

Poor quality and incomplete cerebral ultrasound scans

Overall the reviewers reported one third of the scans to be of poor quality. This is concerning, as these scans had been

Table 6 Variation between the three reviewers (X–Z) in the reported quality and completeness of the cerebral ultrasound scans

| | X | Y | Z | Total |
|--------------------|---------|---------|---------|----------|
| Total babies | 255 | 255 | 255 | 765 |
| Total reports | 255 | 255 | 255 | 765 |
| Incomplete scans | 88 (35) | 93 (37) | 93 (37) | 274 (37) |
| Poor scans | 76 (30) | 93 (36) | 79 (31) | 248 (33) |
| Impossible to read | 10 (4) | 11 (4) | 19 (7) | 40 (5) |
| Not graded | 9 (4) | 13 (5) | 12 (5) | 34 (4) |
| Missing scans | 4 (2) | 4 (2) | 4 (2) | 12 (2) |

Values are number (%) of reports.

Table 7 Variation between the three reviewers (X–Z) and the Australian and New Zealand Neonatal Network (ANZNN) in the reported grades of germinal matrix/intraventricular haemorrhage

| | X | ANZNN Y | ANZNN Z | ANZNN |
|---------------|----------|----------|----------|----------|
| Total babies | 255 | 255 | 255 | 255 |
| Total reports | 232 | 232 | 227 | 220 |
| Normal | 125 (54) | 119 (52) | 118 (52) | 117 (52) |
| Grade 1 | 29 (13) | 32 (13) | 8 (4) | 31 (13) |
| Grade 2 | 11 (5) | 23 (10) | 31 (9) | 22 (10) |
| Grade 3 | 26 (11) | 25 (11) | 16 (7) | 25 (11) |
| Grade 4 | 41 (17) | 33 (14) | 54 (23) | 32 (14) |

Values are number (%) of reports. Scans (n = 86) that were reported as impossible to read, not graded, or missing have been excluded.

Arch Dis Child Fetal Neonatal Ed: first published as 10.1136/adc.2004.065219 on 21 October 2005. Downloaded from <http://fn.bmj.com/> on January 22, 2022 by guest. Protected by copyright.

Table 8 Weighted kappa (κ (w)) statistics showing the level of agreement between the three reviewers (X–Z) and the Australian and New Zealand Neonatal Network (ANZNN) in the reported grades of germinal matrix/intraventricular haemorrhage

| | X | Y | Z |
|--------------|--------------|--------------|--------------|
| Reviewer X | | | |
| κ (w) | 1 | | |
| CI (95%) | | | |
| Reviewer Y | | | |
| κ (w) | 0.84 | 1 | |
| CI (95%) | 0.76 to 0.93 | | |
| Reviewer Z | | | |
| κ (w) | 0.91 | 0.83 | 1 |
| CI (95%) | 0.84 to 0.97 | 0.74 to 0.92 | |
| ANZNN | | | |
| κ (w) | 0.82 | 0.75 | 0.86 |
| CI (95%) | 0.73 to 0.92 | 0.64 to 0.86 | 0.77 to 0.94 |

Values are numbers and 95% confidence intervals (CI). Scans ($n = 86$) that were reported as impossible to read, not graded, or missing have been excluded.

previously reported on in clinical practice, and the reliability of cerebral ultrasound interpretation is reported to be higher when the scan is of high quality.^{14–15} However, we cannot tell from this study if these images were of poor quality at the time the scan was performed or if the poor quality is related to the method of storage.

The methods of storage included digital images, x ray film, and thermal paper. We found that digital images were much clearer than other forms of storage. NICU D, which used a digital system, had no scans reported as impossible to read, and only 4% reported as of poor quality. The images on the thermal paper were never as clear as either digital or x ray film, and faded over time. The earliest scans we used were five years old before being copied for the study. NICUs E and F, which stored scans on thermal paper, or a combination of thermal paper and x ray film, had the highest number of poor quality scans (70%).

These NICUs also accounted for 90% of the incomplete scans. One of these NICUs had no radiology support for cerebral ultrasound scanning, which may have contributed to this finding. If the clinician performing the cerebral ultrasound scan has had limited formal training, they may only take the images of the baby's brain that they know how to interpret. Furthermore, there is no opportunity for collaboration between the neonatal and radiology experts in this NICU. Other recent studies have reported an increase in both the level of correct interpretation and the level of agreement between reviewers when radiology and neonatal staff have collaborated.^{16–17}

Despite this considerable variation in the quality and completeness of the scans between the NICUs, no NICU stood apart from the others as having a greater degree of variation between the reviewers' reports and those to the ANZNN, suggesting that these quality issues did not contribute to the reported variation in incidence of GM/IVH.

Grading of cerebral ultrasound scans

The variation between the NICUs and the reviewers in the reported grade of GM/IVH was small, and the level of agreement for each NICU ranged from good to excellent (κ (w) = 0.61 to 0.99). Agreement was highest for normal scans, and the greatest variations were in grades 1, 3, and 4. Others have also reported greater disagreement between reviewers for grade 1 GM/IVH,^{15–18} perhaps because it can be difficult to distinguish between germinal matrix congestion and haemorrhage.

What is already known on this topic

- The reported incidence of germinal matrix/intraventricular haemorrhage in preterm babies varies between neonatal intensive care units
- Case mix does not explain the variation between units within the Australian and New Zealand Neonatal Network. Other possible explanations include variations in clinical practice, or in the techniques and interpretation of the diagnostic cerebral ultrasonograms

What this study adds

- Variation in the reported incidence of germinal matrix/intraventricular haemorrhage between New Zealand neonatal intensive care units is unlikely to be due to the differences in the capture, storage, and interpretation of the cerebral ultrasonograms
- Further investigation into practice variations is warranted

Grade 3 was both under-reported and over-reported, perhaps because of difficulties in definitions, as was also suggested from the Vermont-Oxford Network.¹⁶ Papile defined grade 3 GM/IVH as a ventricle that is distended with the blood that is within it. However, the ventricle is often distended with both blood and cerebrospinal fluid, and this is included in the definition of a grade 3 GM/IVH in some NICUs.

All three reviewers reported a greater number of grade 4 GM/IVH than had been reported to the ANZNN. This is of concern as this important outcome is sometimes used as a clinical indicator,

Intraobserver and interobserver variability

To measure the level of agreement between reviewers, we used κ (w) statistics, which compares the observed agreement with that expected due to chance.¹⁹ We found excellent agreement between both the original and the re-read reports for each reviewer and between reviewers

Four previous studies have assessed the accuracy of reporting of cerebral ultrasound scans in preterm babies.^{14–16–17–20} All reported that disagreement was common among reviewers. Others have assessed intraobserver and interobserver variability for other types of radiological images, and again describe poorer levels of agreement than in our study, along with statistically significant differences between reviewers.^{21–23} Thus our study is unique in finding a relatively high level of agreement between reviewers across images from different NICUs.

Data collection sheet

We developed and adapted the data collection sheet used for this study from a previously published version.²⁴ Its use may have contributed to the high levels of agreement in our study, as the reviewers were required to report the scan in a uniform way, answering simple observational questions but not determining the grade of GM/IVH. Pinto-Martin and colleagues¹⁷ assessed observation and interpretation separately, and found greater agreement between reviewers when they were asked only to report observations. We are not aware of other published studies that have used this approach to

convert observational data into Papile's grading system. We suggest that future studies should also consider separating observation from interpretation in this way. Introducing a similar data collection sheet into clinical practice may also improve the consistency of reporting of cerebral ultrasound scans.

Summary

A number of studies have reported significant variation in the incidence of GM/IVH between NICUs.^{2,3,5,8} However, in none of these studies did the researchers re-read the original scans, so they could not determine how much of this may be due to variations in recording and reporting of the scans themselves.

In this study we have shown that the quality and completeness of cerebral ultrasound scanning does vary widely between the NICUs in New Zealand. However, despite this, there was a high level of agreement between the reviewers and the reports to the ANZNN, with high intraobserver and interobserver agreement.

We conclude that the reported variation in the incidence of GM/IVH between NICUs in New Zealand is unlikely to be due to the differences in cerebral ultrasound scanning and interpretation. Thus further investigation is warranted into practice variations that may contribute to the differing incidence of GM/IVH in different NICUs.

ACKNOWLEDGEMENTS

This work was supported by the Maurice and Phyllis Paykel Trust, the Rebecca Roberts scholarship, and the Waikato Sick Babies Trust. The ANZNN Advisory Committee and Executive members were: Australia: Centre for Perinatal Health Services Research (New South Wales): David Henderson-Smart and Deborah Donoghue; Flinders Medical Centre (South Australia): Peter Marshall; John Hunter Hospital (New South Wales): Chris Wake; King Edward Memorial and Princess Margaret Hospitals (Western Australia): Noel French, Ron Hagan, and Karen Simmer; Launceston General Hospital (Tasmania): Chris Bailey; Liverpool Health Service (New South Wales): Robert Guarani; Mater Mother's Hospital (Queensland): David Tudehope; Mercy Hospital for Women (Victoria): Andrew Watkins; Monash Medical Centre (Victoria): Kaye Bawden, Andrew Ramsden, and Victor Yu; National Perinatal Statistics Unit (New South Wales): Paul Lancaster; Nepean Hospital (New South Wales): Lyn Downe; Newborn Emergency Transport Service (Victoria): Michael Stewart; New South Wales Newborn and Pediatric Emergency Transport Service: Andrew Berry; Perinatal Research Centre (Queensland): Paul Colditz; Royal Children's Hospital (Victoria): Linda Johnstone and Peter McDougall; Royal Darwin Hospital (Northern Territory): Charles Kilburn; Royal Hobart Hospital (Tasmania): Peter Dargaville; Royal Hospital for Women (New South Wales): Kei Lui; Royal North Shore Hospital (New South Wales): Jennifer Bowen; Royal Prince Alfred Hospital (New South Wales): Nick Evans; Royal Women's Hospital (Queensland): David Cartwright; Royal Women's Hospital (Victoria): Lex Doyle, Colin Morley, and Neil Roy; Sydney Children's Hospital (New South Wales): Barry Duffy; Canberra Hospital (Australian Capital Territory): Graham Reynolds; Children's Hospital at Westmead (New South Wales): Robert Halliday; Townsville Hospital (Queensland): John Whitehall; Western Australia Neonatal Transport Service: Jenni Sokol; Westmead Hospital (New South Wales): William Tarnow-Mordi; Women's and Children's Hospital (South Australia): Ross Haslam; New Zealand: Christchurch Women's Hospital: Nicola Austin; Christchurch School of Medicine: Brian Darlow; Dunedin Hospital: Roland Broadbent; Gisborne Hospital: Graeme Lear; Hastings Hospital: Jenny Corban; Hutt Hospital: Robyn Shaw; Middlemore Hospital: Lindsay Mildenhall; National Women's Hospital: Carl Kushell; Nelson Hospital: Peter McIlroy; Palmerston North Hospital: Jeff Brown; Rotorua Hospital: Stephen Bradley; Southland Hospital: Paul Tomlinson; Taranaki Hospital: John Doran; Tauranga Hospital: Hugh Lees; Timaru Hospital: Philip Morrison; University of Auckland: Jane Harding; Waikato Hospital: David Bouchier; Wairau Hospital: Ken Dawson; Wanganui Hospital: John Goldsmith; Wellington Women's Hospital: Vaughan Richardson; Whakatane Hospital: Chris Moyes; Whangarei Hospital: Peter Jankowitz.

Authors' affiliations

D L Harris, Newborn Intensive Care Unit, Health Waikato Private Bag 3200, Hamilton, New Zealand

D L Harris, F H Bloomfield, J E Harding, Liggins Institute, University of Auckland, Private Bag 92019, Auckland, New Zealand

R L Teele, Department of Paediatric Radiology, Starship Children's Health, Private Bag 92024, Auckland, New Zealand

F H Bloomfield, J E Harding, Newborn Services, National Women's Hospital, Private Bag 92189, Auckland, New Zealand

Competing interests: none declared

REFERENCES

- 1 **Volpe J.** *Neurology of the newborn*, 4th ed. Philadelphia: WB Saunders, 2001.
- 2 **Marteinez A, Taeusch HW, Yu VY, et al.** Variations in mortality and intraventricular haemorrhage in occupants of Pacific Rim nurseries. *J Paediatr Child Health* 2002;**38**:235–40.
- 3 **International Neonatal Network.** Risk adjusted and population based studies of outcome for high risk infants in Scotland and Australia. *Arch Dis Child Fetal Neonatal Ed* 2000;**82**:F118–23.
- 4 **The United Kingdom Neonatal staffing study group.** Patient volume, staffing and workload in relation to risk-adjusted outcomes in a random stratified sample of UK neonatal intensive care units: a prospective evaluation. *Lancet* 2002;**359**:99–107.
- 5 **Synnes AR, Chien LY, Peliowski A, et al.** Variation in intraventricular hemorrhage incidence rates among Canadian neonatal intensive care units. *J Paediatr* 2001;**138**:525–31.
- 6 **International Neonatal Network.** The CRIB score: a tool for assessing initial neonatal risk and comparing performance of neonatal intensive care units. *Lancet* 1993;**342**:193–8.
- 7 **Papile L, Burstein J, Burstein R.** Incidence and evolution of subependymal and intraventricular haemorrhage: a study of infants with birth weights less than 1500 g. *J Paediatr* 1978;**92**:529–34.
- 8 **Heuchan AM, Evans N, Henderson-Smart DJ, et al.** Perinatal risk factors for major intraventricular haemorrhages in Australia and New Zealand Neonatal Network. *Arch Dis Child Fetal Neonatal Ed* 2002;**86**:F86–90.
- 9 **Phibbs CS, Bronstein JM, Buxton E, et al.** The effects of patient volume and level of care at the hospital of birth on neonatal mortality. *JAMA* 1996;**276**:1054–9.
- 10 **Simpson JM, Evans N, Gibberd RW, et al.** Analysing differences in clinical outcomes between hospitals. *Qual Saf Health Care* 2003;**12**:257–62.
- 11 **Donoghue D,** the Australian and New Zealand Neonatal Network. *Annual report of the Australian and New Zealand Neonatal Network.* Sydney: New South Wales Pregnancy and Newborn Services Network: University of Sydney, 2000.
- 12 **Australasian Society for Ultrasound in Medicine.** Guidelines for neurosonography in infants (July 1997). http://www.asum.com.au/open/P&S/policie_d13.htm (accessed 8 June 2005).
- 13 **Lee W.** *Standards for performance of the ultrasound examination of the infant brain.* Laurel: American Institute of Ultrasound Medicine, 1999.
- 14 **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**:273–84.
- 15 **Taylor G, Seibert J, DiPietro M, et al.** Effect of local versus central reading and image quality on sonographic diagnosis of intracranial hemorrhage. In: *American Pediatric Society and European Society for Paediatric Research.* Washington DC: 1996.
- 16 **Carteaux P, Cohen H, Check J, et al.** Evaluation and development of potentially better practices for the prevention of brain hemorrhage and ischemic brain injury in very low birth weight infants. *Pediatrics* 2003;**111**:489–96.
- 17 **Pinto-Martin J, Paneth N, Kazam E, et al.** Interobserver variability in neonatal cranial ultrasonography. *Paediatr Perinat Epidemiol* 1988;**2**:43–58.
- 18 **O'Shea MT, Volberg F, Dillard RG.** Reliability of interpretation of cranial ultrasound examinations of very low-birth weight neonates. *Dev Med Child Neurol* 1993;**35**:97–100.
- 19 **Fleiss JL.** *Statistical methods for rates and proportions.* New York: John Wiley & Sons, 1981.
- 20 **Reynolds PR, Dale RC, Cowan FM.** Neonatal cranial ultrasound interpretation: a clinical audit. *Arch Dis Child Fetal Neonatal Ed* 2000;**84**:F92–5.
- 21 **Rehan VK, Seshia MM, Johnston B, et al.** Observer variability in interpretation of abdominal radiographs of infants with suspected necrotising enterocolitis. *Clin Paediatr* 1999;**38**:637–43.
- 22 **Bloomfield FH, Teele RL, Voss M, et al.** Inter- and intra-observer variability in the assessment of atelectasis and consolidation in neonatal chest radiographs. *Pediatr Radiol* 1999;**29**:459–62.
- 23 **Weatherburn GC, Ridout D, Strickland N. H, et al.** A comparison of conventional film, CR hard copy and PACS soft copy images of the chest: analyses of ROC curves and inter-observer agreement. *Eur J Radiol* 2003;**47**:206–14.
- 24 **Kuban K, Teele RL.** Rationale for grading intracranial hemorrhage in premature infants. *Pediatrics* 1984;**74**:358–63.