Interobserver Reliability for Identifying Specific Patterns of Placental Injury as Defined by the Amsterdam Classification

Raymond W. Redline, MD; Torstein Vik, MD, PhD; Amy Heerema-McKenney, MD; Ann-Helen Jamtoy, MD; Sanjita Ravishankar, MD; Tuyet Nhung Ton Nu, MD; Christina Vogt, MD, PhD; Pamela Ng, MSc; Karin B. Nelson, MD; Stian Lydersen, PhD; Maryam Oskoui, MD, MSc

● Context.—Placental pathology is an essential tool for understanding neonatal illness. The recent Amsterdam international consensus has standardized criteria and terminology, providing harmonized data for research and clinical care.

Objective.—To evaluate the interobserver reliability of these criteria between pathologists at different levels of experience using digitally scanned slides from placentas in a birth population including a large proportion of normal deliveries.

Design.—This was a secondary analysis of selected placentas from a large case-control study of placental lesions associated with neonatal encephalopathy. Histologic slides from 80 placentas were digitally scanned and blindly evaluated by 6 pathologists. Interobserver reliability was assessed by positive and negative agreement, Fleiss κ, and interrater correlation coefficients.

Results.—Overall agreement on the diagnosis, grading, and staging of acute chorioamnionitis and villitis of unknown etiology was moderate to good for all observers and good to excellent for a subset of 4 observers. Agreement on the diagnosis and subtyping of fetal vascular malperfusion was poor to fair for all observers and fair to moderate for the subset of 4 pathologists. Agreement on accelerated villous maturation was poor.

Conclusions.—This study critically evaluates interobserver reliability for lesions defined by the Amsterdam consensus using scanned images with a low frequency of pathologic lesions. Although reliability was good to excellent for inflammatory lesions, lower reliability for vascular lesions emphasizes the need to more explicitly define the specific histologic features and boundaries for these patterns.

(Arch Pathol Lab Med. 2022;146:372–378; doi: 10.5858/arpa.2020-0753-OA)

A consensus conference held in Amsterdam in 2015 convened expert perinatal pathologists from around the world with the goal of introducing standardization to a group of regionalized and somewhat idiosyncratic prior systems of placental diagnosis and nomenclature. This meeting culminated in the 2016 Amsterdam consensus conference document on placental examination and description, which clearly defines 4 major patterns of placental injury and the individual lesions that characterize them.[1] The so-called Amsterdam system is now widely used and has now become the basis for clinical and research activities in the field. Despite those efforts, standardized terminology, specific diagnostic criteria, and illustrations in a single document can fall short of providing the specific and practical diagnostic details needed for reproducible diagnosis by pathologists with variable levels of experience and training who are working in different parts of the world.

The aim of the present study was to investigate the interobserver reliability of microscopic diagnoses using the Amsterdam consensus system criteria, among 6 pathologists with varying levels of experience and training in a large set of digitally scanned placental images from a population of term and near-term infants with a relatively low prevalence of placental lesions.

MATERIALS AND METHODS

Study Population

This is a retrospective secondary analysis of selected placental samples from a case-control study of predictors of neonatal

encephalopathy in term and near-term infants. The original study population consisted of singleton births with gestational age of at least 35 weeks born alive April 1, 2001, to May 31, 2009, from the Royal Victoria Hospital in Montreal.[2] Placental samples were systematically collected and stored during this time period for all inborn infants at this tertiary care center. In the original study placentas of 73 infants with neonatal encephalopathy were compared to 253 randomly selected controls (ratio 1:3). For the purpose of this study, 80 placenta samples from 10 cases and 70 controls were selected based on the quality of available scanned images and enriched for the presence of relevant lesions as recorded in the original study data set by one of the authors.

Placental Samples

Placentas were weighed without membranes on arrival at the Royal Victoria Hospital's Department of Pathology. Placental samples (1–7) were routinely taken from the umbilical cord, placental membrane rolls, and villous parenchyma. All specimens were processed prior to the Amsterdam consensus recommendations, and evaluation of some cases was negatively impacted by the availability of only a single parenchymal section, sometimes from the placental margin and membrane rolls that did not include decidua. The variable quality of membrane rolls precluded assessment of decidual arteriopathy in this study. The blocks were preserved in buffered formalin for at least a day, processed, embedded in paraffin, cut at approximately 5 to 7 microns, and stained with hematoxylin-eosin. Selected slides were then digitally scanned at ×20 for subsequent analysis. Only 1 to 2 relevant scanned images were examined for each case evaluated in the final study.

Study Methods

Scanned placental slides on the 80 study cases were evaluated by 6 observers. These included 2 pathologists with more than 20 years of experience, 3 pathologists with 5 to 20 years of experience, and a pathologist recently out of training. Four of the pathologists work in departments with a much higher annual volume of placental pathology (3000–4000 per year versus <1000). Participants also varied in terms of the types and geographic locations of their perinatal training. All observers had read and implemented criteria from the Amsterdam consensus paper in their practice prior to the study.[1] Before the study started, the observers had a phone conference to review and discuss criteria for all of the variables evaluated (see below). Representative images for a few of the variables were circulated to the group by email. Observers were blinded to study status and clinical history. Gestational age and fetal and placental weights were provided for each case. Scanned images were loaded onto portable hard disks, mailed to the evaluators, and viewed at magnifications up to ×20 using Aperio (Leica Biosystems) or other appropriate digital image computer software as available at each local site. Results were entered into a previously agreed-upon Excel score sheet and sent to the study coordinator for compilation. The compiled files were then forwarded for statistical analyses.

Placental Pathology

The Amsterdam consensus statement provides diagnostic criteria for 4 major patterns of placental injury: 2 inflammatory processes—acute chorioamnionitis (ACA) and villitis of unknown etiology (VUE)—and 2 types of placental vascular insufficiency—fetal vascular malperfusion (FVM) and maternal vascular malperfusion (MVM). Specific findings for these 4 patterns were encoded into 47 separate variables (ACA 17, VUE 9, FVM 14, and MVM 7). In addition to each pathologist having read the original Amsterdam consensus paper, specific written definitions for each variable were provided to each pathologist before scoring. All variables and the number scored positively by each pathologist are listed in Supplemental Table 1 (see the supplemental digital content containing 2 tables at https://meridian.allenpress.com/aplm in the

March 2022 table of contents). Because of the low prevalence of most individual findings in this low-risk cohort, subsequent data analysis was restricted to (1) overall positive impression of ACA, VUE, FVM, and accelerated villous maturation (AVM), a mild component of MVM; (2) grading and staging of the inflammatory responses in ACA and VUE; and (3) selected other important findings present at a high enough prevalence for meaningful interpretation.

Statistics

We calculated specific agreement in terms of negative agreement and positive agreement.[3] Negative agreement and positive agreement have interpretations similar to those of the specificity and sensitivity of diagnostic tests. The 95% CI for specific agreement was calculated as outlined by de Vet et al.[4] We used the Wilson score CI as recommended by Fagerland et al.[5] These specific agreement measures do not show the paradoxes seen with the widely used Cohen κ, and similar agreement measures, such as Scott π and Fleiss κ; that is, low values when the prevalence is low, even if there is high absolute agreement (proportion of patients where the raters agree). However, for the sake of comparisons with other studies, we also report Fleiss κ, which is a generalization of Scott π to more than 2 raters and is similar to Cohen κ. We interpreted Fleiss κ and specific positive and negative agreement in line with the interpretation of κ values proposed by Altman whereby values <0.2 suggest poor agreement, 0.21 to 0.4 fair agreement, 0.41 to 0.6 moderate agreement, 0.61 to 0.8 good agreement, and values >0.80 excellent agreement.[6] Cohen κ was used for pairwise interobserver comparisons and also interpreted using the criteria above. For scale variables, the interrater correlation coefficient was calculated using a 2-way random effects model, with raters and subjects as crossed random factors. In this model, there are 3 variance components: variance between subjects, variance between raters, and the residual variance, with the total variance being the sum of the 3 components. The interrater correlation coefficient is the variance between subjects divided by the total variance and was also interpreted using the criteria outlined above.[7]

Ethics

Institutional ethics board approval was obtained from the McGill University Health Centre (Montreal, Canada) research ethics board prior to initiation of the study. Only de-identified information was employed and data linkage was performed externally to preserve confidentiality.

RESULTS

Selected maternal, neonatal, and gross placental findings for the study population are described in Table 1. Cases from both groups had a low prevalence of complications during pregnancy and at delivery. As expected, placentas from cases with neonatal encephalopathy in the original study (n = 10) were slightly more likely to have an emergency cesarean delivery, meconium-stained amniotic fluid, and 5-minute Apgar score of 4 to 6. They also delivered slightly earlier at a lower birth weight and were more likely to be female than the 70 placentas derived from the matched healthy controls of the original study. None of these differences reached statistical significance.

Overall interobserver agreement for the 4 major patterns of placental injury defined by the Amsterdam classification as estimated by specific positive agreement was good for ACA (0.62; CI, 0.57–0.67) and VUE (0.65; CI, 0.50–0.78), fair for FVM (0.39; CI, 0.28–0.51), and poor for MVM/AVM (0.19; CI, 0.06–0.48). Negative agreement was excellent for all 4 patterns (lowest for FVM). Fleiss κ values were in general slightly lower than the estimates of specific positive agreement (Table 2). Among the 4 observers working in

Table 1. Selected Characteristics of Mothers, and Infants, and Gross Placental Findings of Study Cases		
Neonatal Encephalopathy	**Yes (% Positive)**	**No. (% Positive)**
No.	10	70
Nulliparity	5 (50)[a]	32 (45.7)[a]
Preeclampsia	0	2 (2.9)
Maternal diabetes	0	6 (8.6)
Mode of delivery		
Emergency cesarean	5 (50)	8 (11.4)
Planned cesarean	0	5 (7.1)
Vaginal	5 (50)	57 (81.4)
Meconium in amniotic fluid	2 (25)	12 (17.1)
Gestational age, mean (SD), wk	38.9 (1.60)	39.7 (1.34)
Birth weight, mean, (SD), g	3310 (594)	3500 (452)
Sex		
Male	4 (40)	37 (53)
Female	6 (60)	33 (47)
Apgar score at 5 min		
0–3	0	1 (1)
4–6	1 (10)	0
7–10	9 (90)	69 (99)
Placental weight quartiles[b]		
First: <420 g	3 (30)	14 (20)
Second: 420–489 g	2 (20)	19 (27)
Third: 490–559 g	0	20 (29)
Fourth: ≥560 g	5 (50)	17 (24)
Gross abnormalities of the cord	2 (20)	9 (13)

[a] Column percentage.
[b] Placental weights were categorized into quartiles, based on the values in the control group as detailed in a previous paper.[2]

departments with high annual volumes of placental pathology, positive agreement was improved: good, approaching excellent for ACA (0. 77; CI, 0.57–0.67) and VUE (0.79; CI, 0.71–0.85), and moderate for FVM (0.50; CI, 0.43–0.57), but it remained poor for MVM/AVM (0.19; CI, 0.09–0.39). Representative images for each pattern are provided in the Figure. The implications of these challenges for diagnostic reliability in each category are described in the figure legend and considered at greater length in the Discussion.

The superior reliability of diagnosis for placental inflammatory, as compared with vascular, lesions extended to prognostically important subtypes, including ACA with a stage 2 fetal inflammatory response ($\kappa = 0.73$ for all observers and 0.89 for the subset of 4) and VUE with avascular villi (AV; $\kappa = 0.51$ for all observers and 0.61 for the subset of 4; Table 2 and supplemental digital information, Supplemental Table 2). Furthermore, the ability to effectively scale the severity and extent of inflammation was moderate to good for all observers, with an interrater correlation coefficient of 0.58 for the stage of maternal inflammatory response in ACA, 0.68 for the stage of fetal inflammatory response in ACA, and 0.59 for high-grade versus low-grade features in VUE (Table 3). Stage for the maternal inflammatory response included acute subchorionitis, an obligate precursor of ACA, which was arbitrarily assigned a value of 0.5. Additional analyses excluding so-called stage 0.5 showed similar results (results not shown). Interrater correlation coefficient for the subset of 4 observers improved to the good to excellent range (0.76 for the stage of maternal inflammatory response in ACA, 0.86 for the stage of fetal inflammatory response in ACA, and 0.76 for high-grade versus low-grade features in VUE).

Table 2. Overall Agreement for Entire Group of 6 Examiners and for the Subset of 4 Examiners[a]				
Finding	**Mean Prevalence, %**	**Positive Agreement, Estimate (95% CI)**	**Negative Agreement, Estimate (95% CI)**	**Fleiss κ, Estimate (95% CI)**
All 6 examiners				
ACA, any	11.2	0.62 (0.55–0.69)	0.94 (0.89–0.97)	0.56 (0.43–0.68)
Fetal stage 2	2.0	0.73 (0.66–0.79)	0.99 (0.98–1.00)	0.73 (0.39–1.00)
VUE, any	11.2	0.65 (0.58–0.72)	0.94 (0.90–0.97)	0.59 (0.46–0.73)
VUE, AV	8.3	0.55 (0.48–0.62)	0.96 (0.92–0.98)	0.51 (0.28–0.74)
FVM, any	18.5	0.39 (0.32–0.46)	0.82 (0.75–0.87)	0.21 (0.12–0.29)
AV, small	9.2	0.28 (0.22–0.35)	0.91 (0.86–0.94)	0.19 (0.07–0.31)
AV, large	3.0	0.22 (0.17–0.29)	0.97 (0.93–0.99)	0.19 (0.01–0.37)
VSK, any	2.0	0.23 (0.18–0.30)	0.95 (0.91–0.97)	0.19 (0.06–0.31)
MVM/AVM	3.7	0.19 (0.14–0.25)	0.97 (0.94–0.99)	0.16 (−0.03 to 0.34)
Subgroup of 4 examiners				
ACA, any	13.5	0.77 (0.69–0.83)	0.94 (0.90–0.98)	0.72 (0.58–0.86)
Fetal stage 2	2.3	0.88 (0.83–0.93)	0.997 (0.97–1.00)	0.89 (0.64–1.00)
VUE, any	13.8	0.79 (0.71–0.85)	0.96 (0.91–0.98)	0.74 (0.59–0.90)
VUE, AV	7.5	0.64 (0.53–0.74)	0.96 (0.91–0.98)	0.61 (0.37–0.85)
FVM, any	17.3	0.50 (0.42–0.58)	0.86 (0.80–0.91)	0.37 (0.23–0.50)
AV, small	11.0	0.38 (0.30–0.46)	0.90 (0.84–0.94)	0.28 (0.13–0.43)
AV, large	3.3	0.51 (0.43–0.60)	0.99 (0.94–0.99)	0.49 (0.16–0.83)
VSK, any	4.5	0.44 (0.36–0.53)	0.97 (0.92–0.99)	0.41 (0.10–0.72)
MVM/AVM	3.7	0.20 (0.15–0.28)	0.97 (0.93–0.99)	0.18 (−0.21 to 0.56)

Abbreviations: ACA, acute chorioamnionitis; AV, large, at least 1 large focus of avascular villi; AV, small, exclusively small foci of avascular villi; fetal stage 2, fetal inflammatory response in umbilical artery; FVM, fetal vascular malperfusion; MVM/AVM, maternal vascular malperfusion/accelerated villous maturation; VSK, any, villous stromal vascular karyorrhexis; VUE, villitis of unknown etiology; VUE, AV, VUE with associated avascular villi.

[a] Observers A, B, C, and D (see text and Table 4).
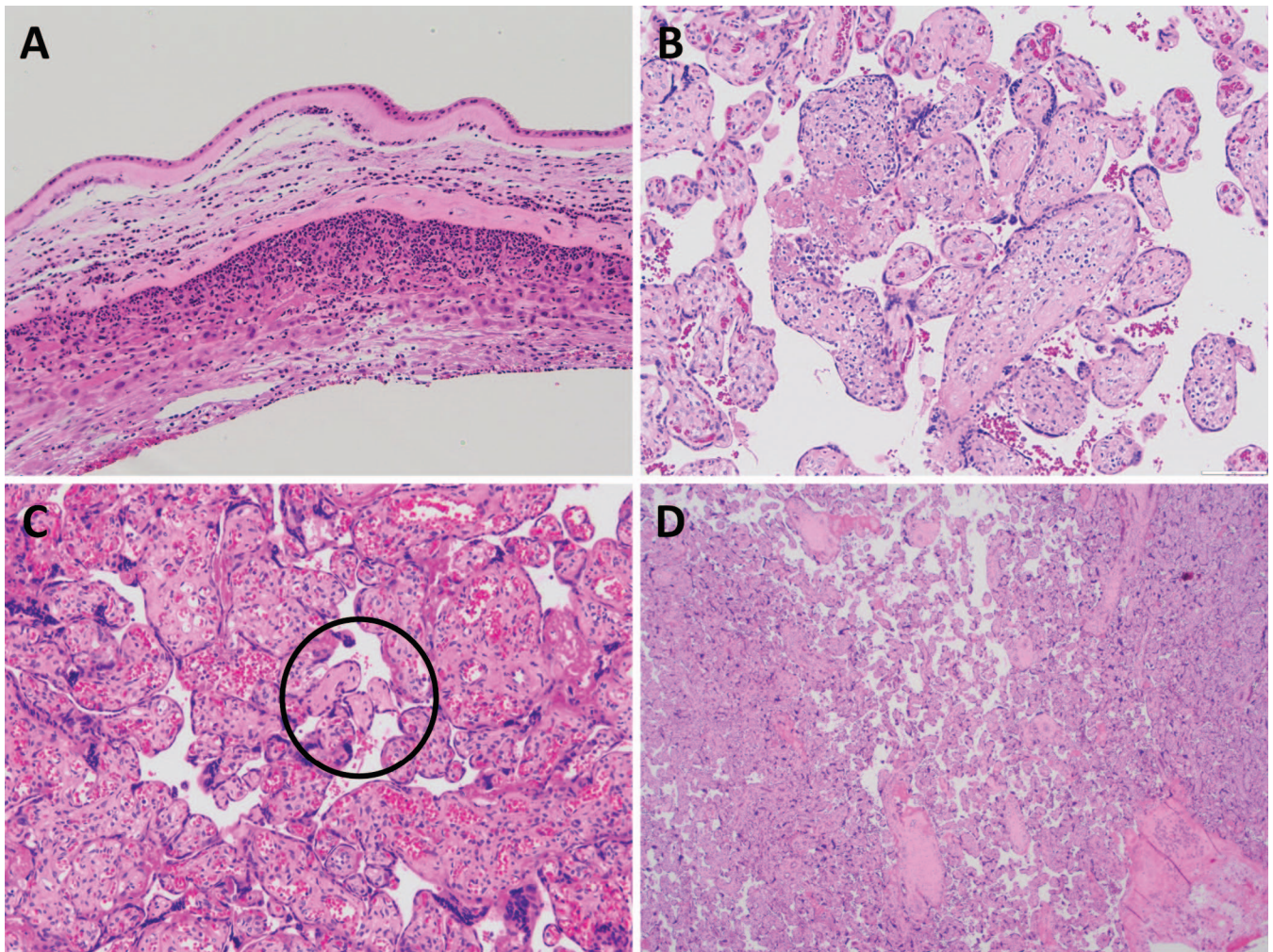
Histologic findings for each of the 4 major patterns of placental injury represent different diagnostic challenges. A, Acute chorioamnionitis, maternal stage 2, placental membranes: a uniform diffuse infiltrate of neutrophils is concentrated within the trophoblastic region of the chorion laeve and spreads into adjacent chorionic and amnionic connective tissue. The presence of this novel exogenous cell population, not seen in the normal placenta, contributes to high diagnostic reliability. B, Villitis of unknown etiology (VUE), low grade. Individual foci of low-grade VUE are by definition small (<10 contiguous villi). However, the patchy asymmetric infiltration of and expansion of the stroma by exogenous lymphocytes, another cell type not seen in the normal placenta, and the accompanying fibrosis and decreased vascularity in comparison with surrounding villi make VUE relatively easy to identify at scanning magnification. C, Fetal vascular malperfusion, small foci of avascular villi (circled): in comparison with VUE, these small foci are composed of fewer (2–4 contiguous villi per focus) and smaller villi. The bland uniform fibrosis of the stroma is less specific, and the lesion can sometimes be confused with re-epithelialized aggregates of intervillous fibrin. Although vessels are absent, the contrast with the vascularity of surrounding villi is less apparent than in VUE. These features all contribute to lower diagnostic reliability. D, Maternal vascular malperfusion, accelerated villous maturation: in contrast to the other lesions illustrated in this figure, accelerated villous maturation in term placentas is most easily recognized at scanning magnification by the observation of alternating areas of villous crowding and paucity. Additional requirements are patchy foci of increased syncytial knotting, villous agglutination, and increased intervillous fibrin within the crowded areas and thin villi with decreased villous branching within the areas of paucity. Low placental weight and increased fetoplacental weight ratio are typical, but they are not by themselves specific diagnostic criteria (hematoxylin-eosin, original magnifications ×10 [A and B], ×20 [C], and ×2 [D]).

Table 3. Interrater Correlation Coefficient Regarding Scaling of Placental Inflammation: Stages of Maternal and Fetal Inflammatory Responses in Acute Chorioamnionitis and Grade in Villitis of Unknown Etiology

Interrater Correlation Coefficient	All 6 Examiners, Estimate (95% CI)	Four Examiners—A, B, C, and D— Estimate (95% CI)
Acute chorioamnionitis		
Maternal inflammatory response, stages 0.5 to 3[a]	0.58 (0.49–0.68)	0.76 (0.68–0.83)
Fetal inflammatory response, stages 1 to 3	0.68 (0.60–0.76)	0.86 (0.82–0.90)
Villitis of unknown etiology, low to high grade	0.59 (0.50–0.69)	0.76 (0.69–0.83)

[a] Stage 0.5 = early acute subchorionitis, a precursor lesion to acute chorioamnionitis, not part of the official staging system.

Table 4. Cohen κ for Pairwise Comparison Between Examiners for Lesions Related to Fetal and Maternal Vascular Malperfusion

Examiners	No.	Four Examiners				Two Examiners	
		A	B	C	D	E	F
Fetal vascular malperfusion, any							
A	17	x	0.27	0.38	0.31	0.21	−0.07
B	13		x	0.41	0.40	0.16	0.15
C	21			x	0.43	0.06	0.07
D	18				x	0.38	−0.02
E	23					x	0.04
F	19						x
Avascular villi, small							
A	13	x	0.05	0.37	0.48	0.27	−0.02
B	14		x	0.04	0.08	0.10	−0.02
C	17			x	0.38	0.04	0.12
D	11				x	0.26	0.13
E	6					x	−0.02
F	6						x
Avascular villi, large							
A	17	x	0.38	0.48	0.49	−0.03	−0.03
B	13		x	0.65	0.38	−0.04	−0.03
C	21			x	0.48	0.05	−0.04
D	18				x	−0.03	−0.03
E	23					x	−0.02
F	19						x
Villous stromal-vascular karyorrhexis, any							
A	3	x	0.65	0.38	0.58	−0.05	−0.05
B	6		x	0.22	0.41	0.10	0.12
C	2			x	0.19	0.22	−0.04
D	7				x	0.08	0.08
E	6					x	0.12
F	5						x
Maternal vascular malperfusion/accelerated villous maturation							
A	5	x	0.32	0.18	0.36	0.18	NA
B	1		x	−0.02	0.32	0.39	NA
C	4			x	0.18	−0.05	NA
D	5				x	0.18	NA
E	4					x	NA
F	0						x

Abbreviation: NA, not applicable.

[a] κ cannot be calculated because examiner F did not select any (N = 0).

Positive agreement for individual lesions within the FVM category was only fair (0.22–0.28), and it was even lower when assessed by Fleiss κ (all values ≤0.21; Table 2). In the subgroup of 4, positive agreement was moderate for 2 of the individual lesions—large foci of AV and villous stromal-vascular karyorrhexis (VSK), 0.51 and 0.44 respectively—whereas it was only fair for small foci of AV (0.38). Again, when assessed with Fleiss κ, agreement was poorer. Further breakdown of the agreement between pairs of individual observers, as estimated by Cohen κ, showed greater overall agreement among the subgroup of 4 than for the other 2 observers (Table 4). Within the subgroup of 4, some distinctions emerged. For large foci of AV, pairwise agreement was fair to good (κ values ranging from 0.38 to 0.65) for all 4 observers. For VSK, agreement was fair to good (κ values ranging from 0.38 to 0.65) for only 3 of the observers (A, B, and D), whereas for small foci of AV, a different subset of 3 (A, C, and D) achieved fair to moderate pairwise κ values (ranging from 0.37 to 0.48).

Estimating reliability for MVM as assessed in this study by AVM was limited by low prevalence (mean, 3.2%), variability (prevalence range 0 to 6%), and the absence of accompanying lesions seen in more severe cases of MVM (ie, villous infarction, infarction-hematoma, and distal villous hypoplasia; see supplemental digital content, Supplemental Table 1). Interobserver agreement was poor (<0.20) for both the group as a whole and the subgroup of 4. However, individual pairwise comparisons in Table 4 did show a fair level of agreement (κ values ranging from 0.32 to 0.36) for 3 of the 6 observers (A, B, and D).

DISCUSSION

The primary goal of this study was to evaluate interobserver reliability between a diverse group of pathologists for identifying the 4 major patterns of microscopic placental lesions defined by the Amsterdam consensus statement. In contrast to most previous studies (discussed below), pathologists evaluated a large cohort of placentas from relatively low-risk term and near-term deliveries in a single round using only definitions provided by the recent Amsterdam document. The Montreal neonatal encephalopathy study cohort[2] was particularly appropriate for this study because it (1) closely approximates the challenges of achieving reliability in actual day-to-day practice and (2) provided the opportunity to use scanned images, which are becoming more widely used in diagnostic pathology and which have the potential to incorporate placental pathology in multicenter clinical and research studies.

Although new classification systems can eventually bring clarity to previously suboptimal systems, the use of unfamiliar terms, variable interpretation of written definitions, and persistent use by some observers of previous criteria may hinder their early application. Variations in experience, training, and specific exposure to a large volume of placental pathology among observers also can play an important role. All of these factors likely contributed to some of the areas of low interobserver reliability found in this study.

Acute chorioamnionitis represents an acute inflammatory response by the mother and fetus to bacteria or fungi in the amniotic fluid.[8] Specific positive agreement between all 6 observers for the overall diagnosis of ACA was good, and it improved for the subgroup of 4 observers practicing in settings with a higher annual volume of placentas submitted to pathology. Kappa values for all 6 observers (0.56) and the subgroup of 4 (0.72) compare favorably with previous studies with a more favorable study design for the examiners (κ values ranging from 0.57 to 0.79).[9–13] Umbilical arteritis (fetal inflammatory response, stage 2) has been recognized in several studies to represent a critical threshold correlated with elevated levels of circulating fetal cytokines.[14,15] In our study, overall agreement for this finding was good (κ = 0.73), comparable to previous studies,[9] and improved to excellent (κ = 0.89) among the subgroup of 4. Although overall assessment of maternal and fetal inflammatory responses in ACA has been assessed in prior studies, detailed evaluation of agreement in staging has not been previously measured. Interobserver correlation coefficients for staging were good overall (maternal 0.58 and fetal 0.68)

and were better for the subgroup of 4 (maternal 0.76 and fetal 0.86). The consistently good reliability seen in multiple studies for diagnosing ACA may be related to specific and straightforward criteria for recognizing neutrophils, cells not normally found in the placenta (Figure, A). Given the variable sensitivity for detecting or accepting very mild inflammatory infiltrates (eg, acute subchorionitis versus an early stage 1 maternal inflammatory response confined to the chorionic plate), this level of reliability is unlikely to be surpassed by further refinement of criteria.

Villitis of unknown etiology is a relatively common lesion characterized by the trafficking of maternal lymphocytes across the villous trophoblast, with subsequent activation and expansion in the fetal villous stroma.[16] It is currently believed that most cases represent alloreactivity to fetal major histocompatibility antigen complex antigens.[17] However, it is acknowledged that a small percentage may represent a reaction to undiagnosed microbial antigens. Reliability of diagnosis for VUE has been understudied. One study reported interobserver and intraobserver agreements of 81% and 85%, respectively, for 3 pathologists after 2 rounds of scoring a selected cohort of 50 slides, 18 of which were considered true positives.[18] Further statistical analysis of these results was not reported. In our study, we found moderate agreement among all observers with a κ of 0.59, with agreement improving to good (κ = 0.74) for the subgroup of 4. A subtype of VUE, cases with AV, has been shown to be a risk factor for fetal central nervous system injury.[19] Reliability of diagnosis for this particular subtype was likewise moderate (κ = 0.51) for the entire group, improving to good (κ = 0.61) for the subgroup of 4. Formal grading of VUE based on the number of contiguous affected villi was introduced in the Amsterdam consensus, and diagnostic agreement has not been previously studied. This simple numeric system for enumerating clusters of inflamed villi is facilitated by their distinction from surrounding normal villi even at relatively low power, as illustrated in Figure, B. Interobserver correlation coefficients for staging were good overall (0.59) and improved for the subgroup of 4 (0.76).

Fetal vascular malperfusion, like MVM, encompasses a spectrum of placental adaptations to decreased perfusion.[20] In the case of FVM, decreased perfusion usually starts with obstruction to flow in the umbilical vein, often due to extrinsic compression or deformation owing to intrinsic cord abnormalities. The resulting stasis and increased venous pressure can lead to a wide spectrum of fetal stromal-vascular alterations, the most specific of which are large foci of AV and VSK. Smaller foci of AV are more difficult to detect (Figure, C), especially on scanned slides, and involve more subjectivity in interpretation. Another problem relating to digital images in this study was that slides were scanned at ×20 rather than ×40 magnification. Although overall agreement among the group of 6 pathologists for FVM was poor, the subgroup of 4 achieved moderate agreement for both large foci of AV and VSK and fair agreement for small foci of AV. This compares favorably with one other study that found moderate agreement for any-sized focus of AV and for VSK in a highly selected population,[21] and is superior to another study that showed essentially no reliability (κ = −0.03) for fetal vascular lesions as a group.[12] Reproducible diagnosis of FVM as a whole and each individual component remains a challenge. Improvement likely depends on more explicit criteria for individual lesions and better delineation of the minimum threshold

and specific combinations of findings required to make an overall diagnosis.

Finally, to understand the low reproducibility for the MVM category, 2 factors need to be taken into account: (1) as noted in the Amsterdam consensus statement, "accelerated villous maturation may be difficult to recognize in a term placenta,"[1] and (2) MVM in our low-risk cohort was largely restricted to the mildest subgroup, those with AVM alone—more severe vascular lesions, such as villous infarction, infarction-hematoma, and distal villous hypoplasia, were not seen in any of the cases by most observers. The absence of moderate to severe cases of MVM and changes in nomenclature make it difficult to compare our results to those of other studies in the literature. Three previous studies evaluated increased syncytial knots, a lesion that overlaps with AVM. One study with a higher prevalence of moderate to severe MVM found moderate agreement.[22] Another that analyzed abnormal placentas with a variety of different lesions found poor to fair agreement,[9] and a third that included both normal and abnormal placentas was unable to calculate agreement because of one observer making the diagnosis of increased syncytial knots in every case.[12] Clearly, reproducible diagnosis of mild MVM in term and near-term placentas at the threshold with normal is a significant challenge for placental pathologists. In our view, the keys to improvement are (1) to recognize the sometimes subtle, alternating areas of crowding and paucity on low-power examination (Figure, D), (2) to better define the specific diagnostic features within these alternating areas that can be seen at higher magnification (ie, clustering of syncytial knots, fibrin and villous agglutination near stem villi in the crowded areas, and decreased caliber and branching of distal villi in the areas of paucity), and (3) to take into account birth weight, gross placental measurements, and possibly clinical history. It is well established that fetal weight and gross measurements (decreased overall weight and an increased fetoplacental weight ratio) are important considerations for the diagnosis of AVM.[23–25] Although this information was provided, it is not clear to what extent it was used by different observers in their evaluations, and this may actually have contributed to low interobserver agreement. Finally, it is well known that AVM is primarily seen in a limited set of clinical situations, including hypertension, fetal growth restriction, and unexplained preterm delivery.[24–27] In our study, the slides were examined without clinical information. Although this is a methodologic requirement of high-quality reliability studies, it may be that for mild AVM at the threshold with normal, pathologists require such information to make a reproducible diagnosis. It is important to distinguish reliability studies from placental evaluation in the real world, where pathologists approach diagnosis with a Bayesian prior probability based on local factors, clinical history, outcomes, and gross pathology, and have the ability to obtain additional history and examine additional sections.

Strengths of our study include the use of standardized scanned images, assessment of every lesion described for the 4 major patterns of placental injury, and a realistic scenario where placentas with a low frequency of abnormalities are examined in a single round by a diverse group of pathologists using Amsterdam consensus definitions. Limitations include the low prevalence of some important lesions, lack of prior discussion of the threshold for positive findings, variable image quality with scanning performed at a maximum of ×20 magnification, preventing examination

at high power (×40), and the fatigue-factor introduced by examining a large number of images in an unfamiliar format (scanned, rather than glass hematoxylin-eosin slides). Despite previous studies in biopsy pathology demonstrating reliability using digital images scanned at ×20 magnification,[28] these last 2 limitations should not be underestimated in the context of much larger placental images. Several observers explicitly noted that cues they normally use to identify lesions on hematoxylin-eosin slides were simply not obvious on the scanned images. Whether this resulted from variability in the quality of the scanned images, the software used for viewing images, or the size and resolution available on individual monitors is not clear, but this is definitely an issue that needs to be addressed in future studies with this study design. Future studies should also address comparisons of reliability on scanned placental images versus the corresponding hematoxylin-eosin–stained slides, the latter of which were unavailable to us in the current study.

In conclusion, our results substantiate and extend those of previous studies showing good to excellent reliability in the evaluation of inflammatory lesions where exogenous cells infiltrate normal placental structures and are easily quantified. Given the variable sensitivity for detecting or accepting very mild inflammatory infiltrates, this level of reliability is unlikely to be improved by further refinement of criteria. Nonetheless, it is noteworthy that our results suggest that the reliability of detecting inflammatory lesions seems to depend in part on experience in terms of volume of placentas examined. Vascular lesions are different in kind, representing a continuum of more subtle changes that change normal structure in several different placental regions. Diagnostic reliability continues to be a challenge for these lesions, especially at the mild end of the spectrum. As suggested by others, participants in future reliability studies may be able to achieve greater agreement by reviewing teaching sets of representative cases together prior to evaluating study cases, and by focusing on smaller blocks of study cases at any given time to avoid the fatigue factor.[29,30] The results of this study suggest that for vascular lesions further refinements and educational efforts to fully implement the Amsterdam system are required to realize its full potential.

References

1. Khong TY, Mooney EE, Ariel I, et al. Sampling and definitions of placental lesions: Amsterdam Placental Workshop Group consensus statement. *Arch Pathol Lab Med*. 2016;140(7):698–713.

2. Vik T, Redline R, Nelson KB, et al. The placenta in neonatal encephalopathy: a case-control study. *J Pediatr*. 2018;202:77–85.e3.

3. de Vet HC, Mokkink LB, Terwee CB, Hoekstra OS, Knol DL. Clinicians are right not to like Cohen's kappa. *BMJ*. 2013;346:f2125.

4. de Vet HCW, Dikmans RE, Eekhout I. Specific agreement on dichotomous outcomes can be calculated for more than two raters. *J Clin Epidemiol*. 2017;83: 85–89.

5. Fagerland M, Lydersen S, Laake P. *Statistical Analysis of Contingency Tables*. London, UK: Chapman and Hall/CRC; 2017.

6. Altman DG. *Practical Statistics for Medical Research*. London, UK: Chapman and Hall; 1991.

7. Rabe-Hesketh S, Skrondal A. *Multilevel and Longitudinal Modeling using Strata*. 3rd ed. College Station, TX: Strata Press Publication; 2012.

8. Blanc W. Amniotic infection syndrome: pathogenesis, morphology, and significance in circumnatal mortality. *Clin Obstet Gynecol*. 1959;2:705–734.

9. Grether JK, Eaton A, Redline R, Bendon R, Benirschke K, Nelson K. Reliability of placental histology using archived specimens. *Paediat Perinat Epidemiol*. 1999;13(4):489–495.

10. Beebe LA, Cowan LD, Hyde SR, Altshuler G. Methods to improve the reliability of histopathological diagnoses in the placenta. *Paediatr Perinat Epidemiol*. 2000;14(2):172–178.

11. Simmonds M, Jeffery H, Watson G, Russell P. Intraobserver and interobserver variability for the histologic diagnosis of chorioamnionitis. *Am J Obstet Gynecol*. 2004;190(1):152–155.

12. Kramer MS, Chen MF, Roy I, et al. Intra- and interobserver agreement and statistical clustering of placental histopathologic features relevant to preterm birth. *Am J Obstet Gynecol*. 2006;195(6):1674–1679.

13. Redline RW, Faye-Petersen O, Heller D, Qureshi F, Savell V, Vogler C. Amniotic infection syndrome: nosology and reproducibility of placental reaction patterns. *Pediatr Dev Pathol*. 2003;6(5):435–448.

14. Rogers BB, Alexander JM, Head J, McIntire D, Leveno KJ. Umbilical vein interleukin-6 levels correlate with the severity of placental inflammation and gestational age. *Hum Pathol*. 2002;33(3):335–340.

15. Kim CJ, Yoon BH, Romero R, et al. Umbilical arteritis and phlebitis mark different stages of the fetal inflammatory response. *Am J Obstet Gynecol*. 2001; 185(2):496–500.

16. Redline RW. Villitis of unknown etiology: noninfectious chronic villitis in the placenta. *Hum Pathol*. 2007;38(10):1439–1446.

17. Enninga EAL, Leontovich AA, Fedyshyn B, et al. Upregulation of HLA-class I and II in placentas diagnosed with villitis of unknown etiology. *Reprod Sci*. 2020; 27(5):1129–1138.

18. Khong TY, Staples A, Moore L, Byard RW. Observer reliability in assessing villitis of unknown aetiology. *J Clin Pathol*. 1993;46(3):208–210.

19. Redline RW. Severe fetal placental vascular lesions in term infants with neurologic impairment. *Am J Obstet Gynecol*. 2005;192(2):452–457.

20. Redline RW, Ravishankar S. Fetal vascular malperfusion, an update. *APMIS*. 2018;126(7):561–569.

21. Redline RW, Ariel I, Baergen RN, et al. Fetal vascular obstructive lesions: nosology and reproducibility of placental reaction patterns. *Pediatr Dev Pathol*. 2004;7(5):443–452.

22. Redline RW, Boyd T, Campbell V, et al. Maternal vascular underperfusion: nosology and reproducibility of placental reaction patterns. *Pediatr Dev Pathol*. 2004;7(3):237–249.

23. Naeye RL. Do placental weights have clinical significance? *Hum Pathol*. 1987;18(4):387–391.

24. Zhou YY, Ravishankar S, Luo G, Redline RW. Predictors of high grade and other clinically significant placental findings by indication for submission in singleton placentas from term births. *Pediatr Dev Pathol*. 2020;23(4):274–284.

25. Redline RW, Patterson P. Patterns of placental injury: correlations with gestational age, placental weight, and clinical diagnosis. *Arch Pathol Lab Med*. 1994;118(7):698–701.

26. Wright E, Audette MC, Ye XY, et al. Maternal vascular malperfusion and adverse perinatal outcomes in low-risk nulliparous women. *Obstet Gynecol*. 2017;130(5):1112–1120.

27. Arias F, Victorio A, Cho K, Kraus FT. Placental histology and clinical characteristics of patients with preterm premature rupture of membranes. *Obstet Gynecol*. 1997;89(2):265–271.

28. Vyas NS, Markow M, Prieto-Granada C, et al. Comparing whole slide digital images versus traditional glass slides in the detection of common microscopic features seen in dermatitis. *J Pathol Inform* 2016;7:30. doi: 10.4103/2153-3539.186909

29. LiVolsi VA. Can we agree to disagree? *Hum Pathol*. 2003;34(11):1081–1082.

30. Morris JA. Information and observer disagreement in histopathology. *Histopathology*. 1994;25(2):123–128.