see commentary on page 477

The Oxford classification of IgA nephropathy: pathology definitions, correlations, and reproducibility

A Working Group of the International IgA Nephropathy Network and the Renal Pathology Society: Ian S.D. Roberts[1], H. Terence Cook[2], Stéphan Troyanov[3], Charles E. Alpers[4], Alessandro Amore[5], Jonathan Barratt[6], Francois Berthoux[7], Stephen Bonsib[8], Jan A. Bruijn[9], Daniel C. Cattran[10], Rosanna Coppo[5], Vivette D'Agati[11], Giuseppe D'Amico[12], Steven Emancipator[13], Francesco Emma[14], John Feehally[6], Franco Ferrario[15], Fernando C. Fervenza[16], Sandrine Florquin[17], Agnes Fogo[18], Colin C. Geddes[19], Hermann-Josef Groene[20], Mark Haas[21], Andrew M. Herzenberg[22], Prue A. Hill[23], Ronald J. Hogg[24], Stephen I. Hsu[25], J. Charles Jennette[26], Kensuke Joh[27], Bruce A. Julian[28], Tetsuya Kawamura[29], Fernand M. Lai[30], Lei-Shi Li[31], Philip K.T. Li[32], Zhi-Hong Liu[31], Bruce Mackinnon[19], Sergio Mezzano[33], F. Paolo Schena[34], Yasuhiko Tomino[35], Patrick D. Walker[36], Haiyan Wang[37], Jan J. Weening[38], Nori Yoshikawa[39] and Hong Zhang[37,*]

Pathological classifications in current use for the assessment of glomerular disease have been typically opinion-based and built on the expert assumptions of renal pathologists about lesions historically thought to be relevant to prognosis. Here we develop a unique approach for the pathological classification of a glomerular disease, IgA nephropathy, in which renal pathologists first undertook extensive iterative work to define pathologic variables with acceptable inter-observer reproducibility. Where groups of such features closely correlated, variables were further selected on the basis of least susceptibility to sampling error and ease of scoring in routine practice. This process identified six pathologic variables that could then be used to interrogate prognostic significance independent of the clinical data in IgA nephropathy (described in the accompanying article). These variables were (1) mesangial cellularity score; percentage of glomeruli showing (2) segmental sclerosis, (3) endocapillary hypercellularity, or (4) cellular/fibrocellular crescents; (5) percentage of interstitial fibrosis/tubular atrophy; and finally (6) arteriosclerosis score. Results for interobserver reproducibility of individual pathological features are likely applicable to other glomerulonephritides, but it is not known if the correlations between variables depend on the specific type of glomerular pathobiology. Variables identified in this study withstood rigorous pathology review and statistical testing and we recommend that they become a necessary part of pathology reports for IgA nephropathy. Our methodology, translating a strong evidence-based dataset into a working format, is a model for developing classifications of other types of renal disease.

Kidney International (2009) **76,** 546–556; doi:10.1038/ki.2009.168; published online 1 July 2009

KEYWORDS: glomerulonephritis; IgA nephropathy; Oxford classification; pathology; renal failure

Correspondence: *Ian S.D. Roberts, Department of Cellular Pathology, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK.*
E-mail: ian.roberts@orh.nhs.uk

*Authors' affiliations are listed in the Acknowledgements.

Received 17 November 2008; revised 24 February 2009; accepted 24 March 2009; published online 1 July 2009

The histological diagnosis of IgA nephropathy is straightforward; it is defined by the presence of IgA-dominant or co-dominant immune deposits within glomeruli, as shown by immunohistochemistry or immunofluorescence. However, biopsies meeting this criterion may show a wide range of histological changes that reflect the clinical diversity of IgA nephropathy. Biopsy appearances may range from virtually normal histology by light microscopy to severe necrotizing, crescentic glomerulonephritis or advanced glomerulosclerosis, and tubular atrophy. There have been numerous clinicopathological studies of IgA nephropathy, the great majority being retrospective, correlating histological changes in diagnostic biopsy with clinical outcome. A number of histological lesions have been reported to be of prognostic value (Table 1).[1–15] The apparently conflicting results of these studies reflect differences in patient cohort, treatment, and clinical outcome measures. In general, studies in which the clinical end point is time to dialysis/renal failure have shown that chronic lesions (tubular atrophy, interstitial fibrosis, and glomerulosclerosis) are the most powerful histological predictors of outcome. This is not surprising, as these lesions reflect an advanced stage of disease; those patients who are biopsied and diagnosed late in the course of their disease will

Table 1 | Histological risk factors for progressive renal failure in IgA nephropathy

Reference	Mesangial cellularity	Endocapillary proliferation	Crescents	Capillary wall IgA	Focal segmental sclerosis	Glomerulosclerosis	Interstitial fibrosis/tubular atrophy
Nozawa *et al.*[1]							X
Ballardie *et al.*[2]	X						
To *et al.*[3]						X	
Mera *et al.*[4]							X
Daniel *et al.*[5]							X
Vleming *et al.*[6]							X
Freese *et al.*[7]			X	X			X
Hogg *et al.*[8]			X			X	
Katafuchi *et al.*[9]					X		X
Ibels *et al.*[10]					X	X	
Okada *et al.*[11]						X	X
Bogenschutz *et al.*[12]							X
Rekola *et al.*[13]	X						
D'Amico *et al.*[14]		X		X		X	
Boyce *et al.*[15]			X				

X, statistically significant association with clinical outcome.

have a shorter time to end-stage renal disease. In contrast, those studies that have correlated histological changes with rate of loss of renal function or response to immunosuppressive therapy have shown that active glomerular lesions (mesangial, endocapillary or extracapillary proliferation, necrosis) are the most significant pathological prognostic factors.

There have been a number of attempts to incorporate the various histological lesions into a pathological classification of IgA nephropathy. None has achieved widespread acceptance. Deficiencies include a lack of definitions and use of vague terminology, lack of an evidence base, and inclusion of both active and chronic lesions in the definition of single categories. For example, a recent classification divides biopsies into four categories (I–IV), namely, no, slight ($<10\%$), moderate (10–30%), and severe ($>30\%$) glomerulosclerosis, crescent formation, or adhesion.[16] Although such a schema may accurately identify those patients who will develop renal failure, it cannot be used to guide patient management; clearly, the management of patients with class IV disease due to $>30\%$ glomerular crescents will differ from those with class IV disease due to diffuse glomerulosclerosis.

As described in the accompanying paper[17], we sought to develop an international consensus classification of IgA nephropathy with a strong evidence base. In this paper, we describe in detail the process by which histological data were collected and reviewed, and present the evidence used for selecting those pathological lesions that were included in the final schema (the 'Oxford Classification' of IgA Nephropathy). The overall philosophy was to collect a highly detailed initial pathological data set and to simplify this into a working schema. We recognize that a 'successful' classification must have clear definitions, be simple to use in routine clinical practice, be reproducible, and have a value independent of the clinical parameters at the time of biopsy. These criteria, therefore, formed the basis of our selection of which lesions to include in the final classification.

RESULTS

Pathology definitions

An initial meeting of pathologists was held in Oxford, UK, in 2005 to define the pathological variables to be assessed in renal biopsies in cases of IgA nephropathy. After a provisional analysis of the first 40 cases, areas of high interobserver variation were identified. To improve reproducibility, the definitions were refined at a meeting of pathologists in Atlanta, USA, in 2006 (Table 2). These were subsequently used for histological scoring of the entire study group. A minor amendment (in italics in Table 2) for defining necrosis in routine practice was agreed upon at a further meeting in Oxford in 2008.

Scoring of histological lesions

A detailed pathology data set was collected initially, with the intention of working to simplify it for use in routine practice. Histology slides from each case were circulated among five pathologists in batches of five, in a rolling manner, to ensure that no two batches were scored by the same five pathologists. A score sheet was completed by individual pathologists for each biopsy (Table 3) using an agreed set of instructions (Table 4). Scoring of mesangial cellularity, together with other proliferative and sclerosing glomerular lesions, is illustrated in Figures 1 and 2. Completed score sheets were collected centrally by one of the pathologists (ISDR) and used to compile the extended pathological data set (Table 5). Completed score sheets were received from five pathologists for 47% of the cases, from four pathologists for 36% of the cases, and from three pathologists for 17% of the cases.

Extended pathology data set

For most histological variables, the median score was taken for analysis (Table 5).

For scoring of glomerular crescents, the mean cellular and fibrocellular crescent scores were obtained by weighing the

Table 2 | Pathological definitions

IgA nephropathy: IgA nephropathy in the native kidney is defined as dominant or codominant staining with IgA in glomeruli by immunofluorescence or immunoperoxidase. Not all glomeruli need show this positivity. SLE-related nephritis should be excluded. The intensity of IgA staining should be more than trace. The distribution of IgA staining should include presence in the mesangium, with or without capillary loop staining, excluding a pure membranous, diffuse, global granular GBM staining pattern or a linear GBM staining pattern. IgG and IgM may be present, but not in greater intensity than IgA, except that IgM may be prominent in sclerotic areas. Complement 3 (C3) may be present. The presence of C1q staining in more than trace intensity should bring up consideration of lupus nephritis.

Glomerular definitions

Diffuse: A lesion involving most ($\geqslant$50%) glomeruli

Focal: A lesion involving $<$50% of glomeruli

Global: A lesion involving more than half of the glomerular tuft (See below for definitions of segmental and global sclerosis)

Segmental: A lesion involving less than half of the glomerular tuft (i.e., at least half of the glomerular tuft is spared). See below for definitions of segmental and global sclerosis

Endocapillary hypercellularity: Hypercellularity due to increased number of cells within glomerular capillary lumina, causing narrowing of the lumina

Karyorrhexis: Presence of apoptotic, pyknotic, and fragmented nuclei

Necrosis is defined by (i) disruption of the glomerular basement membrane with (ii) fibrin exudation and (iii) karyorrhexis. At least two of these three lesions need to be present to meet the criteria for necrosis. (*2008 Amendment: Necrosis should not be scored on the PAS-stained section alone; fibrin is more easily identified on H&E or MSB-stained sections, and breaks in the glomerular basement membrane are more easily identified on silver-stained sections. A minimum requirement for the definition of a necrotizing lesion is extracapillary fibrin exudation.*)

GBM duplication: A double contour of the GBM with or without endocapillary hypercellularity

Increased mesangial matrix: An increase in the extracellular material in the mesangium such that the width of the interspace exceeds two mesangial cell nuclei in at least two glomerular lobules

Sclerosis: Obliteration of the capillary lumen by increased extracellular matrix, with or without hyalinosis or foam cells

An adhesion: An area of continuity between the glomerular tuft and Bowman's capsule separate from an extracapillary lesion or from an area of segmental sclerosis

Segmental sclerosis: Any amount of the tuft involved with sclerosis, but not involving the whole tuft

Global sclerosis: The entire glomerular tuft involved with sclerosis

Collapsed/ischemic glomerulus: A glomerulus showing collapse of the capillary tuft with or without thickening of Bowman's capsule and fibrosis in the Bowman's space

Extracapillary lesions are subclassified as follows:

Extracapillary proliferation or cellular crescent: Extracapillary cell proliferation of more than two cell layers with $>$50% of the lesion occupied by cells. It is further classified by the percentage of glomerular circumference involved: $<$10, 10–25, 26–50, and $>$50%

Extracapillary fibrocellular proliferation or fibrocellular crescent: An extracapillary lesion comprising cells and extracellular matrix, with $<$50% cells and $<$90% matrix. This is further classified by the percentage of the glomerular circumference involved: $<$10, 10–25, 26–50, and $>$50%

Extracapillary fibrosis or fibrous crescent: $>$10% of the circumference of Bowman's capsule covered by a lesion composed of $\geqslant$90% matrix. It is further classified by the percentage of the glomerular circumference involved: 10–25%, 26–50%, and $>$50%. Ischemic, obsolescent glomeruli should be excluded A crescent is one of these extracapillary lesions that involves $>$10% of the circumference of Bowman's capsule

Mesangial hypercellularity is subclassified as follows:

If $<$4 mesangial cells/mesangial area=normal,
4–5 mesangial cells/mesangial area=mild mesangial hypercellularity,
6–7 mesangial cells/mesangial area=moderate mesangial hypercellularity, and
8 or more mesangial cells/mesangial area=severe mesangial hypercellularity.

Note: This is scored for each glomerulus by assessing the most cellular mesangial area. Mesangial areas immediately adjacent to the vascular stalk should not be scored. Individual mesangial areas showing hypercellularity are separated by areas narrowing to the width of $<$2 mesangial cell nuclei (i.e., count clusters, not files of mesangial cell nuclei)

Tubulointerstitial definitions

Tubular atrophy is defined by thick irregular tubular basement membranes with decreased diameter of tubules. It is scored according to the percentage of cortical area involvement, with 1–5% rounded to 5% and other values rounded to the closest 10%

Interstitial fibrosis is defined as increased extracellular matrix separating tubules in the cortical area. It is scored as percentage involvement, with 1–5% rounded to 5% and other values rounded to the closest 10%.

Interstitial inflammation is defined as inflammatory cells within the *cortical* interstitium in excess. It is scored as percentage involvement, with 1–5% rounded to 5% and other values rounded to the closest 10%. It should be noted whether the inflammation is confined to the areas of interstitial fibrosis or not

Additional tubular lesions are noted as follows: The presence of numerous red blood cells, defined as tubules completely filled with red blood cells with or without casts, is noted as a lesion when it involves $\geqslant$20% of tubules

Acute tubular injury of the proximal tubular epithelium is defined by simplification of the epithelium without tubular basement membrane thickening

Vascular definitions

Arterial lesions are scored based on the most severe lesions. Interlobular and larger arteries are scored separately. An interlobular artery is one surrounded by the cortex; an arcuate artery is one at the corticomedullary junction. Intimal thickening is scored by comparing the thickness of the intima to that of the media in the same segment of vessel. Score the intima variously as normal, and thickened to more or less than the thickness of the media.

Arteriolar hyaline is noted as the proportion of arterioles affected (0, 1–25%, 26–50%, $>$50%).

GBM, glomerular basement membrane; H&E, hematoxylin and eosin stain; MSB, Martius scarlet blue; PAS, periodic acid Schiff; SLE, systemic lupus erythematosus.

Table 3 | Score sheet used for collecting detailed histological data set

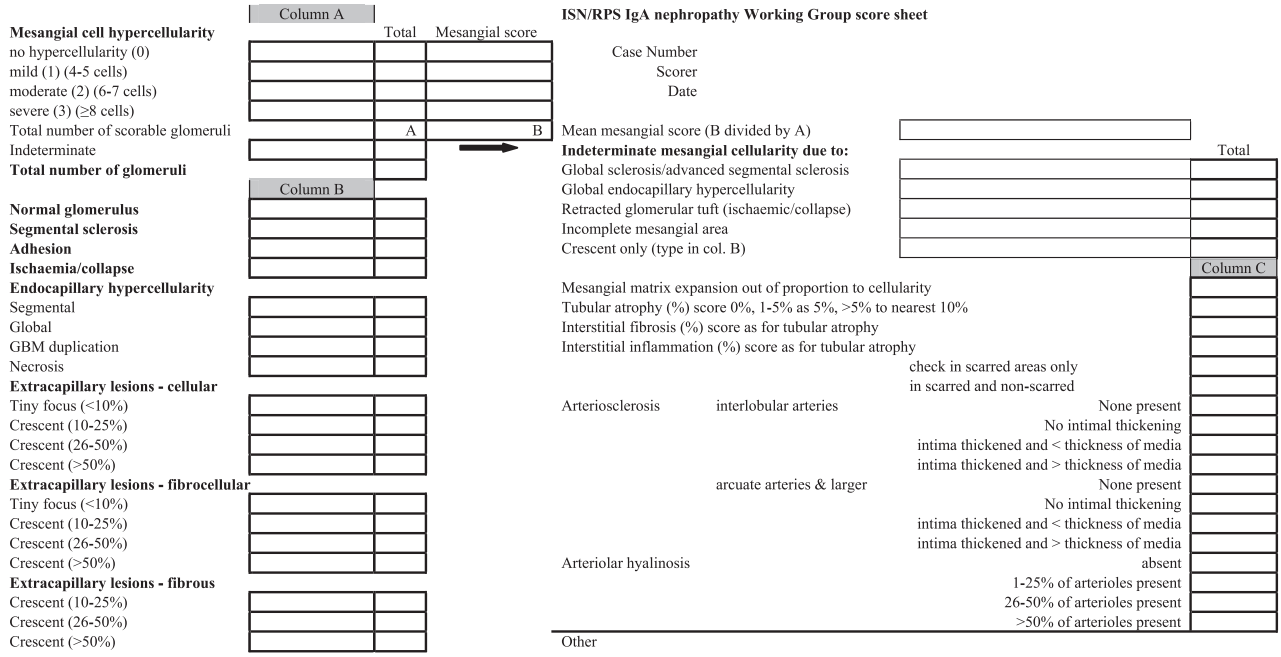

Table 4 | Guidelines for completion of the light microscopy score sheet

1. Using the circled PAS-stained section: for every glomerulus, mark one box only in Column A. If the indeterminate for mesangial cellularity box is marked, then mark which of the five reasons for not scoring mesangial cellularity applies. At least three mesangial cell areas should be present to score a glomerulus. In the global sclerosis category include both solidified and obsolescent glomeruli, and advanced segmental sclerosis when <3 mesangial areas remain. Mesangial cellularity is difficult to score in segments showing endocapillary hypercellularity. Therefore, glomeruli showing global endocapillary hypercellularity should be classed as indeterminate for mesangial cellularity (and the endocapillary lesions noted in column B). Score each glomerulus by assessing the most cellular mesangial area. Mesangial areas immediately adjacent to the vascular stalk should not be scored. Individual mesangial areas showing hypercellularity are separated by areas of narrowing to the width of <2 mesangial cell nuclei (i.e., count clusters, not files, of mesangial cell nuclei). Mesangial cell nuclei are those surrounded by the matrix; do not count those projecting into a capillary lumen
2. Using the circled PAS-stained section: for every glomerulus, mark none, one, or more than one box in Column B as appropriate
   A segmental lesion with capillary occlusion by both sclerosis and endocapillary hypercellularity should be scored for both. Endocapillary hypercellularity is defined by the presence of cells within capillary lumina, not by the matrix. Therefore, in the presence of segmental sclerosis, endocapillary hypercellularity can only be scored within that segment if preserved capillary loops are also present
   GBM duplication: score if it involves an open capillary loop but not as part of a sclerosed segment
3. Using any of the provided sections: for the whole biopsy, mark any box in Column C that applies
   When noting excessive mesangial matrix increase, assess only mesangial areas away from segmental sclerosis, i.e., associated with patent capillary loops
   For scoring arteriolar hyalinosis in Column C, examine only the PAS-stained section used for glomerular scoring
4. In the 'Other' box: note any other abnormality seen, e.g., a glomerular lesion present in one of the sections but not represented in the PAS section used for scoring, mesangiolysis, large numbers of RBC casts, ATN, and malignant vascular disease. Sections should be 2–3 μm thick for scoring. Note if the section appears thicker
5. Total number of glomeruli=total scorable glomeruli+total indeterminate for mesangial cellularity. To produce the mesangial score, multiply the totals of the boxes in column A by 0, 1, 2, or 3 as appropriate. The mean mesangial score is the total of the mesangial scores divided by the number of scorable glomeruli

ATN, acute tubular necrosis; GBM, glomerular basement membrane; PAS, periodic acid Schiff; RBC, red blood cell.

extracapillary lesions by size. A multiplication factor of 1 was applied for lesions $<10\%$ of the glomerular circumference, 2 for lesions 10–25% of the glomerular circumference, 3 for lesions 26–50% of the glomerular circumference, and 4 for lesions $>50\%$ of the glomerular circumference. The resulting scores were summed and divided by the total number of glomeruli in the biopsy.

Additional data items were derived from the completed score sheets to address specific questions. For example,

Mesangial 1 versus 2: Is the proportion of severely hypercellular glomeruli of different significance than that of the mean mesangial cellularity?

Extracapillary 1 versus 2: Are cellular and fibrocellular crescents of different significances?

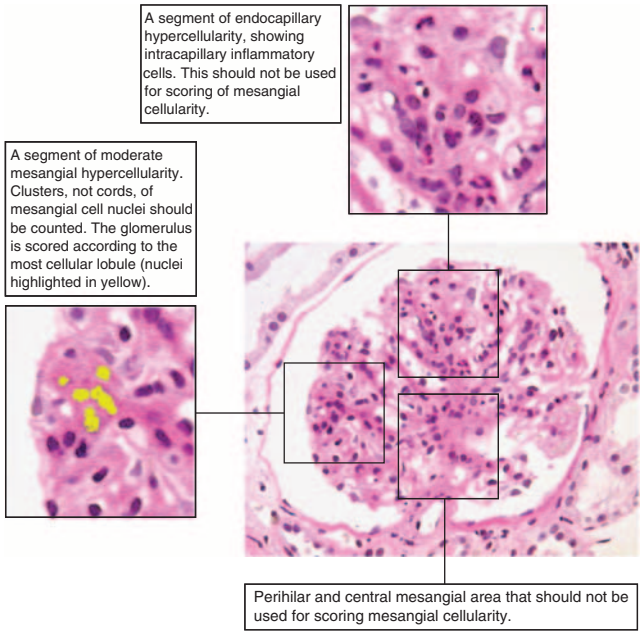

Figure 1 | An illustration of mesangial cellularity scoring (objective ×40 original magnification, periodic acid Schiff stain).

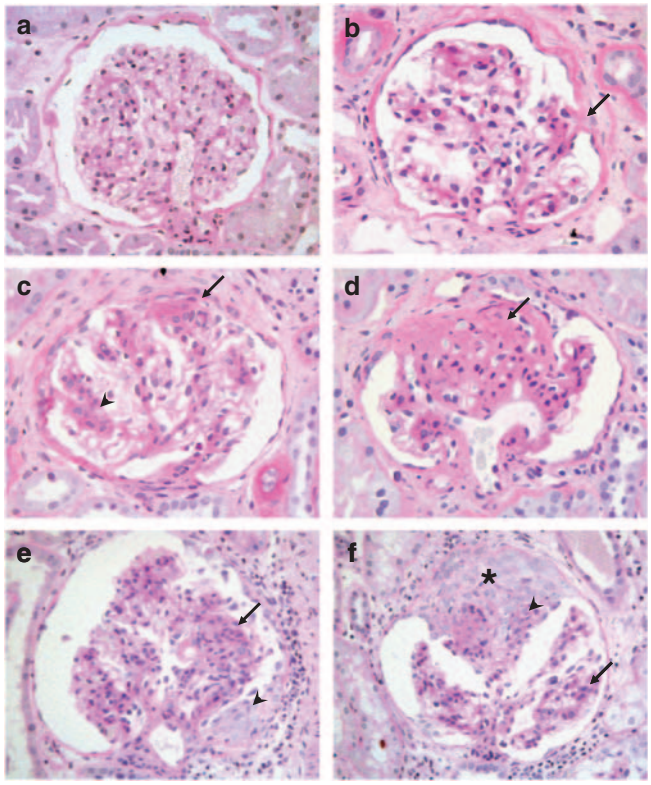

Figure 2 | Proliferative and sclerosing glomerular lesions. All figures objective ×40 original magnification, periodic acid Schiff stain. (**a**) Normal glomerulus by light microscopy. (**b**) Tuft adhesion (arrow) without segmental sclerosis. This lesion should be included with segmental sclerosis lesions for scoring purposes. (**c**) Segmental sclerosis (arrow) with a lobule away from the sclerosis showing moderate mesangial hypercellularity (arrowhead). (**d**) Extensive segmental sclerosis (arrow). This glomerulus should not be used for mesangial scoring. (**e**) A glomerulus showing severe mesangial hypercellularity (arrow) and a small cellular crescent (arrowhead; 10–25% of the glomerular circumference). (**f**) A glomerulus showing mild mesangial hypercellularity (arrow). There is segmental endocapillary hypercellularity (arrowhead); this segment should not be used for mesangial scoring. In addition, there is a cellular crescent (asterisk; 25–50% of the glomerular circumference).

Extracapillary 2 versus 3: Is the size of a crescent significant?

Interstitial inflammation 1 versus 2: Is inflammation confined to areas of fibrosis of different significance compared with inflammation also involving non-fibrotic cortex?

Arteriole 1 versus 2: Is the extent of arteriolar hyalinosis, rather than merely its presence or absence, of significance?

The final, simplified set of pathological variables was selected on the basis of independence from other histological lesions, simplicity of assessment, and reproducibility.

Reproducibility of pathology variables

During the development process, considerable effort was made to minimize interobserver variation between pathologists in the working group. It was agreed that histological lesions that continued to show poor reproducibility within this group should not be a part of the final classification, as the reproducibility is likely to be even lower among pathologists in routine clinical practice. Reproducibility was assessed statistically using intraclass correlation coefficients (ICCs), which are summarized in Table 5.

On the basis of the ICC scores, lesions were divided into three groups as follows:

Group 1: Those lesions showing good or very good reproducibility (>0.6) were mesangial cellularity score, percentage of global glomerulosclerosis, percentage of cellular + fibrocellular crescents, cellular + fibrocellular crescent score (including adjustment for size of crescent), tubular atrophy, interstitial fibrosis, interstitial inflammation 1, and arterial scores 1, 2, and 3.

Group 2: Those lesions showing moderate reproducibility (0.4–0.6) were extent of segmental glomerulosclerosis and percentage of glomeruli showing either segmental or global endocapillary hypercellularity.

Group 3: Those lesions showing poor or fair reproducibility (<0.4) were percentage of normal glomeruli, presence of adhesions, percentage of glomeruli showing segmental endocapillary hypercellularity, presence of glomerular basement membrane duplication, presence of necrosis, percentage of glomeruli showing fibrous crescents, interstitial inflammation 2 (inflammation involving non-fibrotic cortex), and arteriolar hyalinosis.

Those lesions in group 3 were excluded from the classification on the basis of poor reproducibility, with the exception of the following:

- *Adhesions*: Reproducibility increased when combined with segmental sclerosis, indicating that the low ICC score for adhesions alone resulted from different

Table 5 | Extended pathology dataset: definitions and reproducibility

		ICC
Mesangial 1	Median mesangial score	0.64
Mesangial 2	% of scorable glomeruli showing severe mesangial hypercellularity (median of group)	0.54
Global GS	% of total glomeruli showing global sclerosis or retracted glomerular tuft (median of group)	0.90
Normal glomeruli	% of total glomeruli noted as normal (median of group)	0.27
Segmental GS	% of total glomeruli showing segmental sclerosis (median of group)	0.46
Adhesion	% of total glomeruli showing adhesions (median of group)	0.20
Endocapillary 1	% of total glomeruli showing segmental endocapillary hypercellularity (median of group)	0.36
Endocapillary 2	% of total glomeruli showing segmental+global endocapillary hypercellularity (median of group)	0.57
GBM duplication	% of total glomeruli showing GBM duplication (median of group)	0.10
Necrosis	% of total glomeruli showing necrosis (median of group)	0.31
Extracapillary 1	% of total glomeruli showing cellular crescents (median)	0.62
Extracapillary 2	% of total glomeruli showing cellular+fibrocellular crescents (median)	0.64
Extracapillary 3	Mean cellular+fibrocellular crescent score (median of group)	0.66
Extracapillary 4	% of total glomeruli showing fibrous crescents (median)	0.32
Extracapillary 5	Mean fibrous crescent score (median of group)	0.34
Tubular atrophy	% of the cortex showing tubular atrophy (median of group)	0.79
Interstitial fibrosis	% of the cortex showing interstitial fibrosis (median of group)	0.78
Interstitial inflammation 1	% of the cortex showing interstitial inflammation (median of group)	0.58
Interstitial inflammation 2	% of the cortex showing interstitial inflammation if majority (3 or more) checked scarred and non-scarred. Score as 0 if majority checked scarred areas only. Scarred only 0; scarred and non-scarred 1	0.03
Arterial 1	Median arcuate artery score. Leave blank if none present	0.77
Arterial 2	Median interlobular artery score. Leave blank if none present	0.69
Arterial 3	Median artery score—worst of arcuate and interlobular arteries. Leave blank if none present.	0.69
Arteriole 1	Absent=0; present=1. Take majority verdict	0.36
Arteriole 2	Median arteriolar hyalinosis score	0.35

GBM, glomerular basement membrane; GS, glomerulosclerosis; ICC, intraclass correlation coefficient.

pathologists labeling the same lesion as either segmental sclerosis or an adhesion. For subsequent analysis, segmental sclerosis and adhesions were summed.

- *Necrosis*: At the Oxford 2008 meeting, periodic acid schiff (PAS)-stained sections and all other slides from cases in which at least one pathologist had recorded the presence of necrosis were reviewed independently by each of the pathologists. The initial scoring, on which the ICC is calculated, had been carried out on only a single circled PAS-stained section. Reproducibility was higher when all slides, rather than only the PAS-stained slide, were examined (data not shown). This review led to a further refinement of the definition of necrosis to increase reproducibility (see Table 2).
- *Endocapillary hypercellularity*: The ICC for the sum of percentage of segmental and global endocapillary hypercellularity was considerably higher than that for the percentage of segmental hypercellularity, indicating that there was poor reproducibility for the distinction of segmental from global lesions rather than for the identification of endocapillary hypercellularity. The sum of segmental and global endocapillary hypercellularity was, therefore, used in subsequent analyses.

Correlation between pathology variables

Significant correlations between 23 pathology variables (excluding 'normal glomeruli') are shown in Table 6. Given the 253 different comparisons possible, the initial significance was set at $P = 0.05/253$, that is, $P = 0.0002$. Seventy-seven comparisons were considered statistically significant. Several of the strong correlations are not unexpected; for example, the correlations between interstitial fibrosis and tubular atrophy and between both of these and global glomerulosclerosis. However, there are other significant correlations that may be important in terms of pathogenesis. Thus, segmental sclerosis correlates with extracapillary lesions including either fibrocellular or fibrous crescents, suggesting a common pathogenesis. It is also of interest that capillary wall duplication, although poorly reproducible, does show a significant correlation with endocapillary proliferation.

Although many of the pathology variables showed a significant correlation with others, the correlation coefficient between some was so close to 1 that to include both in a classification would provide no additional value. For example, the '*R*' values for endocapillary 1 and 2 (0.99), extracapillary 2 and 3 (0.99), interstitial fibrosis and tubular atrophy (0.98), interstitial fibrosis and interstitial inflammation (0.9), interstitial fibrosis and global glomerulosclerosis (0.8), arterial 2 and 3 (0.9), and arteriole 1 and 2 (0.95) indicated that these pairs of variables were very closely linked.

The selection of which of the linked variables to include in the classification was based on reproducibility, ease of identification, and susceptibility to sampling error. For example, extracapillary 2, a simple calculation of % cellular + fibrocellular crescents, was preferred to extracapillary 3, a complex calculation requiring scoring of the size of the crescents in each glomerulus. Interstitial fibrosis and tubular atrophy were preferred to global glomerulosclerosis,

Table 6 | Correlations between pathology variables

	Mesang		GS			Endocap				Extracap					Interstitium				Vessels				
	Mes1	Mes2	GlobGS	SegGS	Adh	End1	End2	GBMdup	Necr	Extr1	Extr2	Extr3	Extr4	Extr5	TubAt	IntFib	IntInfl1	IntInfl2	Art1	Art2	Art3	Artiol1	Artiol2
Mes1	—	0.7		0.2	0.3	0.3	0.3				0.2	0.2				0.2	0.3						
Mes2		—			0.3	0.3	0.3																
GlobGS			—	0.4	0.2										0.7	0.8	0.6			0.3	0.3	0.4	0.3
SegGS				—	0.5						0.4	0.4	0.3	0.3	0.5	0.5	0.4						
Adh					—						0.3	0.3	0.2	0.2	0.3	0.3	0.3						
End1						—	0.99	0.3		0.4	0.5	0.5											
End2							—	0.3		0.4	0.5	0.5											
GBMdup								—															
Necr									—														
Extr1										—	0.7	0.7											
Extr2											—	0.99	0.4	0.4									
Extr3												—	0.4	0.4			0.2						
Extr4													—	0.99									
Extr5														—									
TubAt															—	0.98	0.9		0.8	0.3	0.3	0.4	0.4
IntFib																—	0.9		0.8	0.3	0.3	0.4	0.4
IntInfl1																	—					0.3	
IntInfl2																		—					
Art1																			—	0.8	0.9	0.6	
Art2																				—	0.9	0.5	0.5
Art3																					—	0.5	0.5
Artiol1																						—	0.95
Artiol2																							—

Adh, adhesion; Art, arterial; Artiol, arteriole; End, endocapillary; Extr, extracapillary; GBMdup, glomerular basement membrane duplication; GlobGS, global glomerulosclerosis; IntFib, interstitial fibrosis; IntInfl, interstitial inflammation; Mes, mesangial; Necr, necrosis; SegGS, segmental glomerulosclerosis; TubAt, tubular atrophy.
Only statistically significant *R* values (correlation coefficients) are shown.
Statistically significant correlations were determined using the Holm–Bonferroni method to minimize the probability of making a Type I statistical error.

as their quantification is less susceptible to error owing to a paucity of glomeruli or because of subcapsular sampling.

Mesangial hypercellularity score

On the basis of our selection criteria, mesangial hypercellularity score was included in the final schema. As reported in the accompanying paper, the optimal cutoff given by sensitivity analysis for predicting clinical outcome was 0.71, which was approximated to 0.5 for clinicopathological correlations, without a significant loss of sensitivity. Mesangial score was derived from scoring each individual glomerulus and taking the mean. Although it is reproducible, simple to perform, and of clinical significance, for some biopsies with large numbers of glomeruli, it can be time consuming. We therefore reasoned that not all pathologists would be willing or have time to score mesangial cellularity formally in routine practice. Therefore, a simpler alternative was tested, that is, dividing biopsies according to whether more or less than half of the glomeruli show mesangial hypercellularity. The relationship between this measure and the mesangial hypercellularity score was formally assessed at the final Oxford meeting. All pathologists were asked independently to provide a percentage of glomeruli showing mesangial hypercellularity for 16 cases from the study group that did not show endocapillary or extracapillary lesions (8 with mesangial score >0.7 and 8 with mesangial score <0.7). For all but one case with an original mesangial score of <0.7, the majority of pathologists scored the biopsies as $<50\%$ of glomeruli showing hypercellularity. For all cases with an original mesangial score of >0.7, the majority of pathologists scored the biopsies as $>50\%$ of glomeruli showing hypercellularity. For cases near the borderline, with an original mesangial score of 0.5–0.7, not surprisingly, there was high interobserver variation for the cutoff of 50% of glomeruli showing hypercellularity. On the basis of this evaluation, it was concluded that dividing biopsies according to whether more or less than 50% of glomeruli show mesangial hypercellularity is a suitable alternative to the formal mesangial hypercellularity score for use in everyday practice. However, for research studies and clinical trials, formal mesangial hypercellularity scores are recommended.

Pathology variables assessed in the final classification

As described above, the initial pathology variables were refined by excluding those with poor interobserver reproducibility and only including one variable from those pairs or groups that were shown to be strongly correlated. This left the following variables, all common in IgA nephropathy, to be further analyzed in relation to the clinical data:

(1) mesangial cellularity score;
(2) percentage of glomeruli showing segmental adhesions or sclerosis;
(3) percentage of glomeruli showing endocapillary hypercellularity;
(4) percentage of glomeruli showing cellular or fibrocellular crescents;
(5) percentage of interstitial fibrosis/tubular atrophy; and
(6) arterial score

The accompanying paper describes the further analysis of these variables in relation to clinical presentation and outcome.

Adequacy of biopsies for classification

The minimum number of glomeruli for a biopsy to be included in the study was initially set at eight. The median number of glomeruli in the 265 biopsies meeting inclusion

criteria was 18. To determine whether the number of glomeruli in a biopsy influences the histological scores, the glomerular number was correlated with scores for the 25 histological lesions. There was no significant correlation, other than with endocapillary and extracapillary proliferation. These showed a weak positive correlation with the number of glomeruli in a biopsy. For endocapillary 2 versus number of glomeruli, Spearman's correlation coefficient was 0.22 ($P \leqslant 0.001$), and for extracapillary 2 versus number of glomeruli, Spearman's correlation coefficient was 0.15 ($P = 0.014$).

To better illustrate these findings, the biopsies were then divided into three groups according to the number of glomeruli: 8–12 ($n = 69$), 13–17 ($n = 59$), and $\geqslant 18$ glomeruli ($n = 137$). Scores for the 25 histological lesions were compared between these groups. There was no significant difference in the mean score for any lesion between the biopsies with 8–12 and 13–17 glomeruli. Biopsies with $\geqslant 18$ glomeruli showed marginally but statistically significant higher mean scores for only three lesions: mesangial 2 (8.1 ± 13.4 versus 5.7 ± 13.2 for 8–12 glomeruli, and 6.0 ± 13.4 for 13–17 glomeruli, $P = 0.01$), endocapillary 1 (6.3 ± 8.9 versus 5.3 ± 11.8 for 8–12 glomeruli, and 5.3 ± 10.3 for 13–17 glomeruli, $P = 0.01$), and endocapillary 2 (7.9 ± 13.1 versus 6.5 ± 5.1 for 8–12 glomeruli and 5.9 ± 12.2 for 13–17 glomeruli, $P = 0.006$).

DISCUSSION

Our aim was to design a systematic approach for the development of a reproducible pathological classification of IgA nephropathy that would predict clinical outcome. To this end, we collected cases with defined clinical outcomes and assessed a range of features in the renal biopsies. We proceeded by first assessing the reproducibility of the scoring of individual biopsy features, then asking which features showed good reproducibility and were independent, and finally asking which of those were related to presenting clinical parameters and had an independent relevance to clinical outcome. Although this seems to be a logical way to develop a classification, this approach has not generally been followed in renal pathology. Thus, the classification schemes for lupus nephritis, as first defined in the World Health Organization (WHO) classifications and subsequently in the International Society of Nephrology/Renal Pathology Society (ISN/RPS) classification,[18] were developed by groups of experts without any attempt to show that the definitions or the classes were reproducible entities. Subsequent testing has shown areas of high interobserver variation, with κ-scores < 0.4 for differentiating active from chronic and segmental from global in class IV disease.[19] In addition, the classifications were not tested for their predictive value before publication, although this has been done subsequently.[20,21] With regard to earlier classification schemes in IgA nephropathy, several of these have been tested for their predictive value, but, in many cases, the variables used in the classifications were not systematically defined, and in no case was there an attempt to examine the reproducibility of the histological variables that were assessed. For example, in the Lee classification,[22] the vague terminology used (terms such as mostly, occasional, frequent, and localized) makes the schema difficult to be employed in a reproducible way. Consequently, earlier histological classifications of IgA nephropathy have not been accepted by the majority of nephropathologists. In a 2006 Renal Pathology Society survey, only 37% of responding pathologists used a classification system for reporting IgA nephropathy biopsies. Five different schema were used, of which that of Haas[23] was the most popular, but even this was used by only 14% of pathologists (38% of those who used a classification system).

The range of histological features we studied was restricted to those that can be assessed by light microscopy, and all the included features had been suggested by earlier studies to have an effect on outcome. Our initial definitions followed those used earlier in the WHO atlas of glomerular diseases[24] and in the classification of lupus nephritis.[18] Although the definitions seemed straightforward and well established, it is important to recognize that some have not been easy to apply in practice. Perhaps the most critical of these was the definition of mesangial hypercellularity, which is a typical feature in IgA nephropathy. Our scoring system depended on assessing the number of nuclei in glomerular mesangial areas, but it became clear at the second meeting of pathologists that there was uncertainty about what constituted a mesangial area; hence, a revision of the definition was necessary. A problem was also encountered with the definition of necrosis, which needed to be revised as shown in Table 2. Cellular crescents are most commonly defined as extracapillary proliferation involving at least 25% of the glomerular circumference. We also noted smaller foci of extracapillary proliferation (10–25 and $< 10\%$ of the glomerular circumference) in our biopsies, and sought to determine whether the size, rather than merely the extent, of crescents was of independent significance. The correlation between the percentage of glomeruli with crescents and the crescent score (that included a multiplier for size of the individual crescentic lesions) was very close indeed, $r = 0.99$, indicating that subdividing crescents by size provided no additional information.

For each of the variables scored, we assessed the ICC. It is notable that some features in which the definitions seemed simple were poorly reproducible, for example, glomerular basement membrane duplication, fibrous crescents, and arteriolar hyaline. We have not further examined why there is such poor agreement on these features. One limitation of our study is that, in order to ensure consistency across pathologists, we restricted the scoring to a single PAS-stained section. Although we believe that most of the features examined would be seen well in PAS-stained sections, this might explain the variability of glomerular basement membrane duplication that is more reliably assessed in silver stains. We also found that necrosis could not be reliably

assessed on the PAS stain. It is also of note that reproducibility of the percentage of 'normal' glomeruli was poor. This may not be surprising when one considers the nature of IgA nephropathy. Unlike pauci-immune vasculitic glomerulonephritis, in which lesions are truly focal and segmental, all glomeruli in IgA nephropathy are abnormal to some extent, in that all contain mesangial IgA deposits. The most minor changes are detectable only on immunostaining or on electron microscopy, but large deposits may be seen on PAS stain, even in the absence of proliferation, and a subtle increase in mesangial matrix frequently accompanies the deposits. It is likely that the difference between pathologists in identifying these very mild changes at light microscopy explains the poor ICC for 'normal' glomeruli.

It could be argued that the good interobserver reproducibility we have achieved for some variables reflects the fact that the scoring was carried out by a group of pathologists who had met together on several occasions and that such good reproducibility may not translate into clinical pathology practice. Although this may be true, we feel that it is important that we have shown that the features retained in the classification had good reproducibility and, when several different features were strongly correlated, we used the one feature easiest to identify and least susceptible to sampling error.

An important question for many renal diseases is the issue of how much tissue is required for reliable diagnosis and classification. The answer depends, to a certain extent, on the nature of the condition. In general, diffuse glomerular diseases will require fewer glomeruli than those in which the pathology is focal. In the case of IgA nephropathy, glomerular IgA deposits are diffuse and a biopsy containing a single glomerulus may be sufficient to make a firm diagnosis. Many of the glomerular lesions, however, are focal, including endocapillary and extracapillary proliferation and segmental sclerosis. Thus, to apply a classification that includes quantitation of these lesions will require more than one glomerulus. We initially set the minimum number of glomeruli for inclusion in the study at eight. This limit was chosen, as it had been used in earlier studies of IgA nephropathy[2] and was similar to the criterion used in other conditions, such as the Banff classification of allograft pathology (8 glomeruli) and the ISN/RPS classification of lupus nephritis (10 glomeruli). We subsequently analyzed the histological scores according to biopsy size, in order to test the validity of using eight glomeruli as a criterion for adequacy. It is not surprising that some of the focal lesions (endocapillary and extracapillary proliferation scores) showed a weak correlation with the number of glomeruli. This is unlikely to be clinically relevant; those biopsies with the fewest glomeruli (8–12) showed no difference in mean scores compared with those with 13–17 glomeruli. Although those biopsies containing numbers of glomeruli above the median for the whole group (18) had significantly more severe mesangial and endocapillary lesions than those below the median, it would be impractical to exclude 50% of biopsies from classification on the basis of a minor difference in some focal lesions.

The statistical methodology used to develop this classification merits clarification. We used the ICC to address the reliability of multiple raters. This flexible method is an extension of the commonly used κ-statistic that is used to assess the agreement between two diagnostic tests, but applied to >2 raters and/or ordinal or continuous measurements.[25] Perfect agreement is indicated by an ICC of 1, and pure chance is indicated by a score of 0. The interpretation of the coefficient levels can be subjective, but authors have suggested a minimum of 0.4 as being necessary to define fair agreement.[26,27]

Given the number of variables studied and the exploratory nature of the work, numerous statistical tests were carried out, but appropriate precautions were taken to minimize type 1 error, that is, falsely rejecting the null hypothesis and assuming statistically significant differences between two variables. This was addressed using the Holm–Bonferroni method for multiple comparisons,[28,29] a valid modification of the more stringent Bonferroni correction.[30]

In summary, we have described here a systematic approach to developing a histological scoring scheme in IgA nephropathy. The results that we have found for interobserver reproducibility for individual features are likely to be applicable to other types of glomerulonephritis, but it should not be assumed that this will be true for the correlations between variables that may depend on the underlying pathobiology of each glomerular disease. We believe that our approach can act as a model for developing classifications for other types of renal disease. The accompanying paper describes the way in which the histological features described here relate to clinical outcome.

MATERIALS AND METHODS

The overall design of the study, patient cohort, clinical data set, and clinicopathological correlations are described in the accompanying paper. Briefly, clinical data and renal biopsy material from 265 patients with IgA nephropathy were collected from 8 countries from 4 continents. Five centers from Asia, six from Europe, two from the United States, one from South America, and two multicenter networks (Canada and the United States) participated.

Biopsies containing <8 glomeruli were regarded as inadequate for scoring and were excluded from the analysis.

Statistical methods

We assessed reproducibility for each variable of the extended pathology data set using ICC.[31] The ICC is a measure of reproducibility applicable to multiple raters. By convention, ICC of <0.40 is poor inter-rater reliability, 0.40–0.59 is moderate, 0.60–0.79 is substantial, and 0.80 is outstanding.[32,33]

Correlations between pathology variables were carried out using the Pearson test or the Spearman test appropriately. Given the number of possible comparisons between pathology variables, we used the Holm–Bonferroni method to minimize the risk of making a type 1 statistical error.[28,29] Briefly, this methodology compares the smallest P-value of all (k) comparisons with an α-value of $0.05/k$. If that P-value is $<0.05/k$, the association is considered to be

statistically significant (reject the null hypothesis). The next smallest *P*-value is then compared with $0.05/(k-1)$, the following with $0.05/(k-2)$, etc. This continues until a *P*-value is superior to the calculated α-value, at which point, the procedure is stopped and all remaining comparisons are considered not statistically different (accept all other null hypotheses).

Analyses were carried out using SPSS software (version 11, SPSS. Chicago IL, USA).

DISCLOSURE

All the authors declared no competing interests.

ACKNOWLEDGMENTS

The work was supported by an unrestricted educational grant from Vifor Aspreva Pharma. The Working Group acknowledges the generous support of the International Society of Nephrology, Kidney Research UK, and Vifor Pharma Aspreva.

[1]Department of Cellular Pathology, John Radcliffe Hospital, Oxford, UK; [2]Imperial College, London, UK; [3]Hôpital du Sacré-Coeur de Montréal, University of Montreal, Montreal, Quebec, Canada; [4]Department of Pathology, University of Washington Medical Center, Seattle, Washington, USA; [5]Nephrology, Dialysis and Transplantation Unit, Regina Margherita Children's Hospital, University of Turin, Turin, Italy; [6]The John Walls Renal Unit, Leicester General Hospital, Leicester, UK; [7]Department of Nephrology, Dialysis, and Renal Transplantation, Hôpital Nord, CHU de Saint-Etienne, Saint-Etienne, France; [8]Department of Pathology, LSU Health Sciences Center, Shreveport, Louisiana, USA; [9]Department of Pathology, Leiden University Medical Center, Leiden, The Netherlands; [10]University Health Network, Toronto General Research Institute, Toronto, Ontario, Canada; [11]Department of Pathology, Columbia University College of Physicians & Surgeons, New York, New York, USA; [12]Fondazione D'Amico per la Ricerca sulle Malattie Renali, Milan, Italy; [13]Department of Pathology, Case Western Reserve University, Cleveland, Ohio, USA; [14]Department of Nephrology and Urology, Division of Nephrology and Dialysis, Bambino Gesù Children's Hospital and Research Institute, Piazza S Onofrio, Rome, Italy; [15]Renal Immunopathology Center, San Carlo Borromeo Hospital, Milan, Italy; [16]Division of Nephrology and Hypertension, Mayo Clinic, Rochester, Minnesota, USA; [17]Department of Pathology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands; [18]Department of Pathology, Vanderbilt University, Nashville, Tennessee, USA; [19]The Renal Unit, Western Infirmary, Glasgow, UK; [20]Department of Cellular & Molecular Pathology, German Cancer Research Center, Heidelberg, Germany; [21]Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; [22]Department of Pathology, University Health Network and University of Toronto, Ontario, Canada; [23]St Vincent's Hospital, Melbourne, Australia; [24]Scott and White Medical Center, Temple, Texas, USA; [25]Division of Nephrology, Hypertension and Renal Transplantation, College of Medicine, University of Florida, Gainesville, Florida, USA; [26]Department of Pathology and Laboratory Medicine, University of North Carolina, Chapel Hill, North Carolina, USA; [27]Division of Immunopathology, Clinical Research Center Chiba, East National Hospital, Chiba, Japan; [28]Department of Medicine, University of Alabama at Birmingham, Birmingham, Alabama, USA; [29]Division of Nephrology and Hypertension, Jikei University School of Medicine, Tokyo, Japan; [30]The Chinese University of Hong Kong, Hong Kong; [31]Research Institute of Nephrology, Jinling Hospital, Nanjing University School of Medicine, Nanjing, China; [32]Department of Medicine, Prince of Wales Hospital, Chinese University of Hong Kong, Hong Kong, China; [33]Departamento de Nefrología, Escuela de Medicina, Universidad Austral, Valdivia, Chile; [34]Renal, Dialysis and Transplant Unit, Policlinico, Bari, Italy; [35]Division of Nephrology, Department of Internal Medicine, Juntendo University School of Medicine, Tokyo, Japan; [36]Nephropathology Associates, Little Rock, Arkansas, USA; [37]Renal Division of Peking University First Hospital, Peking University Institute of Nephrology, Beijing, China; [38]Erasmus Medical Center, Rotterdam, The Netherlands and [39]Department of Pediatrics, Wakayama Medical University, Wakayama City, Japan

REFERENCES

1. Nozawa R, Suzuki J, Takahashi A *et al.* Clinicopathological features and the prognosis of IgA nephropathy in Japanese children on long-term observation. *Clin Nephrol* 2005; **64**: 171–179.
2. Ballardie FW, Roberts IS. Controlled prospective trial of prednisolone and cytotoxics in progressive IgA nephropathy. *J Am Soc Nephrol* 2002; **13**: 142–148.
3. To KF, Choi PC, Szeto CC *et al.* Outcome of IgA nephropathy in adults graded by chronic histological lesions. *Am J Kidney Dis* 2000; **35**: 392–400.
4. Mera J, Uchida S, Nagase M. Clinicopathologic study on prognostic markers in IgA nephropathy. *Nephron* 2000; **84**: 148–157.
5. Daniel L, Saingra Y, Giorgi R *et al.* Tubular lesions determine prognosis of IgA nephropathy. *Am J Kidney Dis* 2000; **35**: 13–20.
6. Vleming LJ, de Fijter JW, Westendorp RG *et al.* Histomorphometric correlates of renal failure in IgA nephropathy. *Clin Nephrol* 1998; **49**: 337–344.
7. Freese P, Norden G, Nyberg G. Morphologic high-risk factors in IgA nephropathy. *Nephron* 1998; **79**: 420–425.
8. Hogg RJ, Silva FG, Wyatt RJ *et al.* Prognostic indicators in children with IgA nephropathy—report of the Southwest Pediatric Nephrology Study Group. *Pediatr Nephrol* 1994; **8**: 15–20.
9. Katafuchi R, Oh Y, Hori K *et al.* An important role of glomerular segmental lesions on progression of IgA nephropathy: a multivariate analysis. *Clin Nephrol* 1994; **41**: 191–198.
10. Ibels LS, Gyory AZ. IgA nephropathy: analysis of the natural history, important factors in the progression of renal disease, and a review of the literature. *Medicine (Baltimore)* 1994; **73**: 79–102.
11. Okada H, Suzuki H, Konishi K *et al.* Histological alterations in renal specimens as indicators of prognosis of IgA nephropathy. *Clin Nephrol* 1992; **37**: 235–238.
12. Bogenschutz O, Bohle A, Batz C *et al.* IgA nephritis: on the importance of morphological and clinical parameters in the long-term prognosis of 239 patients. *Am J Nephrol* 1990; **10**: 137–147.
13. Rekola S, Bergstrand A, Bucht H. IGA nephropathy: a retrospective evaluation of prognostic indices in 176 patients. *Scand J Urol Nephrol* 1989; **23**: 37–50.
14. D'Amico G, Minetti L, Ponticelli C *et al.* Prognostic indicators in idiopathic IgA mesangial nephropathy. *Q J Med* 1986; **59**: 363–378.
15. Boyce NW, Holdsworth SR, Thomson NM *et al.* Clinicopathological associations in mesangial IgA nephropathy. *Am J Nephrol* 1986; **6**: 246–252.
16. Wakai K, Kawamura T, Endoh M *et al.* A scoring system to predict renal outcome in IgA nephropathy: from a nationwide prospective study. *Nephrol Dial Transplant* 2006; **21**: 2800–2808.
17. A Working Group of the International IgA Nephropathy Network and the Renal Pathology Society: Cattran D, Coppo R, Cook T *et al.* The Oxford Classification of IgA nephropathy: Rationale, clinicopathological correlations and classification. *Kidney Int* 2009; **76**: 534–545.
18. Weening JJ, D'Agati VD, Schwartz MM *et al.* The classification of glomerulonephritis in systemic lupus erythematosus revisited. *J Am Soc Nephrol* 2004; **15**: 241–250.
19. Furness PN, Taub N. Interobserver reproducibility and application of the ISN/RPS classification of lupus nephritis—a UK-wide study. *Am J Surg Pathol* 2006; **30**: 1030–1035.
20. Yokoyama H, Wada T, Hara A *et al.* The outcome and a new ISN/RPS 2003 classification of lupus nephritis in Japanese. *Kidney Int* 2004; **66**: 2382–2388.

21. Hill GS, Delahousse M, Nochy D *et al.* Class IV-S versus class IV-G lupus nephritis: clinical and morphologic differences suggesting different pathogenesis. *Kidney Int* 2005; **68**: 2288–2297.
22. Lee SMK, Rao VM, Franklin WA *et al.* IgA nephropathy: morphologic predictors of progressive renal disease. *Hum Pathol* 1982; **13**: 314–322.
23. Haas M. Histologic subclassification of IgA nephropathy: a clinicopathologic study of 244 cases. *Am J Kid Dis* 1997; **29**: 829–842.
24. Churg J, Bernstein J, Glassock RJ. *Renal Disease: Classification and Atlas of Glomerular Diseases*. Igaku-Shoin: New York, 1995.
25. Hripcsak G, Heitjan DF. Measuring agreement in medical informatics reliability studies. *J Biomed Inform* 2002; **35**: 99–110.
26. Fleiss JL. *Statistical Methods for Rates and Proportions*. John Wiley: New York, 1981, pp 212–236.
27. Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics* 1977; **33**: 363–374.
28. Holm S. A simple sequentially rejective multiple test procedure. *Scand J Stat* 1979; **6**: 65–70.
29. Norman G, Streiner D. *Biostatistics, The Bare Essentials*. Hamilton, Ontario: BC Decker Inc, 2000.
30. Aickin M, Gensler H. Adjusting for multiple testing when reporting research results: the Bonferroni vs Holm methods. *Am J Public Health* 1996; **86**: 726–728.
31. Armstrong GD. The intraclass correlation as a measure of interrater reliability of subjective judgments. *Nurs Res* 1981; **30**: 314–315, 320A.
32. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159–174.
33. Koch GG, Landis JR, Freeman JL *et al.* A general methodology for the analysis of experiments with repeated measurement of categorical data. *Biometrics* 1977; **33**: 133–158.