Nephrol Dial Transplant (2008) 23: 223–230
doi:10.1093/ndt/gfm555
Advance Access publication 2 November 2007

Original Article

NDT
Nephrology Dialysis Transplantation

Interobserver agreement of scoring of histopathological characteristics and classification of lupus nephritis

Cecile Grootscholten[1], Ingeborg M. Bajema[2], Sandrine Florquin[3], Eric J. Steenbergen[4], Carine J. Peutz-Kootstra[5], Roel Goldschmeding[6], Marc Bijl[7], E. Christiaan Hagen[8], Hans C. van Houwelingen[9], Ronald H. W. M. Derksen[10] and Jo H. M. Berden[1]

[1]Division of Nephrology, Radboud University Nijmegen Medical Center, Nijmegen, [2]Department of Pathology, Leiden University Medical Center, Leiden, [3]Department of Pathology, Academic Medical Center, Amsterdam, [4]Department of Pathology, Radboud University Nijmegen Medical Center, Nijmegen, [5]Department of Pathology, Academic Hospital Maastricht, Maastricht, [6]Department of Pathology, University Medical Center Utrecht, Utrecht, [7]Division of Clinical Immunology, University Medical Center Groningen, Groningen, [8]Department of Internal Medicine, Meander Medical Center, Amersfoort, [9]Department of Medical Statistics, Leiden University Medical Center, Leiden, and [10]Department of Rheumatology & Clinical Immunology, University Medical Center Utrecht, Utrecht for the Dutch Working party on SLE

Abstract

Background. Assessing renal biopsies from patients with lupus nephritis (LN) is a difficult task and it is subject to interobserver variability. In this study the interobserver agreement amongst five nephropathologists was analysed.

Methods. Five specialized nephropathologists scored 126 biopsies, comprising 87 first and 39 repeat biopsies from 87 patients with biopsy-proven proliferative LN, included in a randomized controlled trial. The interobserver agreement [expressed as intraclass correlation coefficients (ICC)] of the scored histopathological items was calculated. Also, the WHO1995 and ISN/RPS2003 classification systems for LN were compared, with extra attention being given to the comparison between patients with diffuse proliferative LN with either segmental (IV-S) or global (IV-G) lesions.

Results. There was a wide range of agreement. A good interobserver agreement (ICC > 0.6) was present in 15%, and a moderate interobserver agreement (ICC 0.4–0.6) in 31% of the scored items. The activity index for LN showed a good (ICC 0.716) and the chronicity index a moderate (ICC 0.494) interobserver agreement. Both classification systems showed low agreement, although consensus was easily reached. Patients classified as IV-S ($n = 15$) had more favorable clinical parameters at study entry than those with class IV-G ($n = 57$). Although suggested by others, we found no differences in outcome between these two subclasses.

Conclusions. This study shows that, although definitions were agreed upon beforehand, even specialized on nephropathologists have difficulties with scoring histopathological characteristics of LN, particularly with SLE the classification systems.

Keywords: agreement; kidney biopsy; lupus nephritis; observer; systemic lupus erythematosus

Introduction

Systemic lupus erythematosus (SLE) is an autoimmune disease in which virtually any organ can be affected [1]. Involvement of the kidney occurs frequently [2] and the morphological changes observed in the kidney show a great diversity [3]. Attempts have been made to categorize the morphological changes in order to recognize certain patterns, with the aim of providing better guidelines for therapy and prognosis. Until the beginning of this century, one of the WHO classification systems for lupus nephritis (LN) was used, the latest being the WHO1995 classification [3]. In 2003, the classification for LN was redefined by the International Society of Nephrology/Renal Pathology Society (ISN/RPS) [4] in order to allow better standardization of renal biopsies in lupus patients than before [4,5]; in particular as it separates segmental and global lesions. The importance of the latter was suggested by Najafi *et al.* [6] in 2001. Patients with diffuse proliferative LN and segmental lesions ($n = 24$) were more likely to reach end-stage renal disease than patients with diffuse proliferative LN and global lesions ($n = 35$), although the number of patients

Correspondence to: Jo H.M. Berden, MD, PhD, Division of Nephrology (464), Radboud University Nijmegen Medical Center, PO Box 9101, 6500 HB Nijmegen, The Netherlands. Email: j.berden@nier.umcn.nl

entering remission and developing a renal flare was not different.

In a search for a less toxic alternative for cyclophosphamide pulse therapy, the first Dutch LN study started in 1995. In this study, 87 patients with biopsy-proven proliferative LN were randomized to either cyclophosphamide pulse therapy, or to azathioprine combined with methylprednisolone pulses [7]. After 2 years a repeat renal biopsy was performed in 39 participants.

In order to evaluate the histomorphological lesions of these biopsies in detail, a standardized scoring form was developed. Here we report the interobserver agreement among five specialized nephropathologists. We also compare the WHO1995 classification with the ISN/RPS2003 classification for LN.

Subjects and methods

Patient selection

From September 1995 until September 2001, 87 patients with biopsy-proven proliferative LN were included in the first Dutch LN study [7]. In this randomized controlled trial, patients were treated with either azathioprine combined with methylprednisolone and prednisone, or cyclophosphamide pulse therapy with prednisone for 2 years. After 2 years, this regimen was followed by long-term azathioprine and prednisone. All patients were followed regularly, and clinical and laboratory parameters were collected according to the study protocol [7].

Renal biopsy

All biopsies were performed and processed locally. Before randomization, all 87 inclusion biopsies were reviewed and classified by at least one of three experienced nephropathologists according to the WHO1995 criteria for LN [3]. The activity index (AI) and chronicity index (CI) were calculated for each specimen as described previously [8], with maximum scores of 24 for the AI and 12 for the CI. After 2 years, all available patients were asked to have a repeat renal biopsy performed for study purposes only. It was agreed that findings in these biopsies would be anonymized and not used for therapeutic decisions. The ethics committees of all participating hospitals approved this pathology study, and it was conducted according to the Declaration of Helsinki. After informed consent, 41 patients underwent a repeat renal biopsy. For this study silver-stained slides (with HE-background), and for some items PAS slides, were used.

Scoring procedure

Scoring form. Histological characteristics described in LN, together with its definitions, were gathered from the literature. In three consensus meetings with three specialized nephropathologists, the list of items was reduced and a concept scoring form was made. In a pilot study, 10 biopsies were scored to solve discrepancies and problems. In a consensus meeting, these discrepancies were discussed and the scoring form was adjusted. In another consensus meeting, the scoring form was again discussed and revised with five different, experienced nephropathologists (I.B., S.F., E.S., C.P., R.G.) and a specialized nephropathologist from abroad (Dr G. S. Hill, Hôpital Europeen Georges Pompidou, Paris, France). With the final version and after collecting the slides at a central office, three scoring sessions followed. Each biopsy was scored by three of five nephropathologists.

Item selection and definitions. The first pathologist marked the section with the most glomeruli and this slide was subsequently used for the scoring procedure. For almost all assessments, silver-stained slides were used (exceptions were predefined). Scoring the biopsy was divided into several sections: glomerular, tubular, interstitial and vascular. For definitions see our previous publication [9]. For each glomerulus, the presence of a variety of histopathological items was counted. After having scored the section, the nephropathologist was asked to give an opinion on the overall quality of the biopsy: either poor, reasonable or excellent. The quality was evaluated on both the quality of the staining and the quality of the material.

The assessment was completed by filling out the WHO1995 and ISN/RPS2003 classification and activity and chronicity indices for LN [3,4,8]. For these aspects of the assessment, the definitions belonging to the classification systems and the activity and chronicity indices were used. To simplify the comparisons between the two classification methods, membranous features were mentioned as '+V'. After all scoring sessions were finished, the discrepancies in classification were discussed and a final joint judgement was made, during three consensus meetings.

Statistics

All data were analysed by using SPSS 12.0.1 software. All continuous glomerular variables were expressed as the percentage of the total number of non-sclerotic and non-ischemic glomeruli scored in the section.

Intraclass correlation coefficients (ICC) were used to evaluate interobserver readings for the entire set of slides [10]. The ICC is an index of concordance that indicates the degree of agreement beyond that expected by chance alone, and is appropriate when assessing agreement between two or more raters. It is expressed in a score ranging between 0 and 1:0 indicating agreement by chance and 1 total agreement. In general, values higher than 0.8 are considered as excellent, values between 0.6 and 0.8 as good, values between 0.4 and 0.6 as moderate, and values below 0.4 as poor concordance. The ICCs of both continuous, binominal and ordinal variables can be computed, which also takes into account the ranking order of the variables. The ICC was favored over the Cohen Kappa value since each of the five observers scored only 3/5 of the biopsies. The scoring of the biopsies was organized in such a way that an overlap of at least one-third of the slides existed between two nephropathologists.

To account for possible systematic differences between nephropathologists, the nephropathologists were added as 'random effects'. In doing so, the found

ICCs are comparable with the weighted Kappa. Although the classification systems for LN are not actual ordinal variables, for this analysis they were regarded to be ordinal. Comparing the classes without their subcategories would be too rough.

For the descriptive parameters and the comparisons, only those histopathological characteristics with an ICC > 0.4 were used. The mean value of the scores of the three pathologists who scored the biopsy was taken. Non-parametric tests were used for the comparison between patient subgroups. A P-value <0.05 was regarded statistically significant.

Results

Biopsies

In total 128 renal biopsies were performed in 87 patients. All patients underwent a biopsy at study entry. Forty one patients agreed to undergo a repeat biopsy after two years of therapy. Of these 41 repeat biopsies, two did not contain glomeruli and were not included, leaving 126 biopsies for scoring. Overall, the quality of the preparation of the biopsies was good: 55% were judged as excellent and 28% as reasonable. However, in 17% of the biopsies either the staining or cutting quality was judged as poor. A total of nine biopsies could not be classified. Reasons were: six or less glomeruli ($n = 7$; five repeat biopsies and two taken at study entry), or poor quality ($n = 2$, both repeat biopsies).

Interobserver agreement

The analyses were performed for all 126 biopsies. A wide range of agreement for the diverse characteristics was found. Good or excellent concordance (ICC > 0.6) was present in 15% of all 54 scored characteristics. Agreement was moderate (ICC 0.4–06) in 31% and low (ICC < 0.4) in 54% of all scored characteristics. The ICCs did not improve if the analyses were done after exclusion of biopsies with six or less glomeruli or biopsies of poor quality. The observed ICCs also did not change if only the biopsies taken at study entry were analysed. In Table 1 the variables for which excellent, good or moderate agreement existed are listed.

Glomeruli. Excellent or good intraclass correlation coefficients were found for the following items: total number of glomeruli (ICC 0.950), completely sclerosed glomeruli (0.821), endocapillary proliferation (0.646), extracapillary proliferation (0.639) and spikes (0.629). For several items a moderate ICC was found (Table 1).

Tubules and interstitium. Two tubular variables (tubular atrophy and casts) were scored with moderate agreement. Interstitial infiltrate and fibrosis showed acceptable ICCs, while disagreement existed for edema.

Table 1. Interobserver agreement: items for which agreement was excellent (ICC > 0.8), good (ICC 0.6–0.8) or moderate (ICC 0.4–0.6)

Item	ICC
Glomerular	
Total glomeruli	0.950
Completely sclerosed glomeruli	0.821
Endocapillary proliferation[a]	0.646
Extracapillary proliferation[a]	0.639
Spikes[a]	0.629
Synechia[a]	0.522
Wire loops[a]	0.498
Karyorrhexis[a]	0.482
Ischemic glomeruli[a]	0.455
Loop necrosis[a]	0.439
Tubulo-interstitial	
Interstitial infiltrate	0.514
Tubular atrophy	0.511
Casts	0.458
Interstitial fibrosis	0.418
Indices	
Activity index	0.716
Chronicity index	0.494

ICC, intraclass correlation coefficient. For the following items a poor interobserver agreement (ICC < 0.4) was found: double contours, normal glomeruli, mesangiolysis, mesangial proliferation, glomerular infiltration (both mononuclear and polynuclear), mesangial sclerosis (both segmental and global), lumenal macrophages, tubular necrosis, interstitial edema, WHO1995 classification, ISN/RPS2003 classification, and some subscores of the activity and chronicity indices: cellular proliferation, fibrinoid necrosis/karyorrhexis, leukocyte infiltration, tubulointerstitial mononuclear infiltration, glomerular sclerosis, fibrous crescents, interstitial fibrosis. The vascular items were not present often.
[a]Percentage of glomeruli affected by the characteristic that was scored (expressed as percentage of non-sclerotic and non-ischemic glomeruli).

Vascular changes. Vascular changes were not observed very often. Only for arterial fibrous intimal hyperplasia and arterial thrombosis, acceptable agreement was observed (ICC 0.461 and 0.421, respectively).

Activity and chronicity index. Good concordance was present for the AI (ICC 0.716) and moderate concordance for the CI (ICC 0.494).

WHO1995 classification and ISN/RPS2003 classification

Remarkably low ICCs (ICC < 0.2) were found for the classification systems, using both the WHO1995 and the ISN/RPS2003 classification. If for the WHO1995 classification the subclasses were excluded, and the classes II, III, IV, V and VI were compared, the level of agreement slightly improved (ICC improved from 0.108 to 0.368). And, if the classes were divided into class II, III, IV-S, IV-G, V and VI (ISN/RPS2003 classification) the ICC increased from 0.182 to 0.414.

Table 2 shows how, after reaching consensus, all biopsies were classified. Membranous features were present in 6 (25%) and in 8 (10%) of the biopsies classified as class III and IV, respectively.

Table 2. WHO1995 and ISN/RPS2003 classification in 117 biopsies (85 biopsies at study entry and 32 repeat biopsies)

WHO1995	*n*	ISN/RPS2003	*n*
I	0		
IIa	11	II	10
IIb	3	II	1
IIIa	5	III-A	5
IIIb	7	III-A/C	7
IIIc	12	III-C	12
IVa	4	IV-G-A	4
IVb	13		
		IV-S-A	1
		IV-S-A/C	1
		IV-G-A	6
		IV-G-A/C	5
IVc	56		
		IV-S-A/C	15
		IV-G-A	2
		IV-G-A/C	39
IVd	5		
		IV-S-C	2
		IV-G-C	3
V	0		3
VI	1	VI	1
total			117

Three biopsies were classified as class II LN in the WHO1995 classification, but due to the presence of membranous features these were classified as class V in the ISN/RPS2003 classification

Segmental vs *global diffuse proliferative LN*

In view of the assumed prognostic importance, patients with ISN/RPS2003 class IV-S and IV-G were assessed separately. At study entry, 15 biopsies were classified as class IV-S and 57 as class IV-G. The clinical characteristics are summarized in Table 3. Patients with segmental lesions had more favorable characteristics: their renal function was better, hypertension was less often present, and complement values and hemoglobin were higher. A number of outcome parameters were not different between the two groups. Actually, more patients had doubling of serum creatinine in the IV-G group ($n=5$) than in the IV-S group ($n=0$), after a median follow-up of 77 months (interquartile range (IQR) 54–96), but the differences were not statistically significant. Log rank analysis of (time to) partial and complete remission did not show differences between the two groups. Not unexpectedly, more endocapillary proliferation and higher scores for wire loops were present in the biopsies categorized as class IV-G. Also, the total activity index was lower in class IV-S (7.7 *vs* 9.7), but this difference was not significant (Table 4 gives the histological characteristics).

To allow a comparison between the ISN/RPS2003 and the WHO1995 classification, the clinical and

Table 3. Clinical characteristics of 72 patients with class IV lupus nephritis at study entry and during follow-up.

	IV-S $n=15$	IV-G $n=57$	*P*	IVa or IVb $n=17$	IVc or IVd $n=55$	*P*
Clinical characteristics						
CY/AZA (n)	9/6	35/22	ns	7/10	37/18	ns
Age (year)	37 (28–47)	31 (24–41)	ns	26 (21–34)	33 (24–44)	0.037
Female (*n*) (%)	12 (80)	50 (88)	ns	12 (71)	50 (91)	0.034
Caucasian (*n*) (%)	12 (87)	40 (70)	ns	12 (71)	41 (75)	ns
Nephritis in past (*n*) (%)	3 (20)	12 (21)	ns	1 (6)	14 (26)	ns
Hypertension (*n*) (%)	4 (27)	35 (61)	0.017	8 (47)	31 (56)	ns
Systolic blood pressure (mmHg)	140 (120–145)	140 (120–150)	ns	140 (120–153)	140 (120–150)	ns
Diastolic blood pressure (mmHg)	80 (70–90)	80 (80–90)	ns	80 (78–98)	80 (80–90)	ns
Disease duration (months)	2.5 (0.4–109)	22 (1–82)	ns	2 (1–27)	22 (1–102)	ns
SLEDAI	20 (14–26)	21 (16–24)	ns	22 (17–27)	20 (14–24)	ns
Laboratory parameters						
Screat (μmol/l)[a]	95 (72–121)	112 (87–154)	0.039	98 (84–139)	115 (85–152)	ns
Proteinuria (g/24 h)	3.5 (1.8–6.5)	4.2 (2.7–6.6)	ns	3.9 (2.8–6.2)	4.1 (2.2–6.7)	ns
Hemoglobin (mmol/l)	7.6 (7.0–8.3)	6.2 (5.6–7.2)	0.000	6.9 (5.6–7.6)	6.3 (5.6–7.5)	ns
C3 (g/l)	0.54 (0.47–0.77)	0.43 (0.32–0.60)	0.012	0.45 (0.24–0.62)	0.49 (0.38–0.63)	ns
C4 (g/l)	0.12 (0.10–0.22)	0.10 (0.07–0.12)	0.008	0.08 (0.07–0.14)	0.10 (0.08–0.14)	ns
Anti-dsDNA (IU/ml)	141 (37–558)	166 (18–553)	ns	342 (93–807)	124 (22–502)	ns
Outcome parameters						
Follow-up (months)	78 (63–95)	76 (54–95)	ns	93 (54–107)	75 (54–88)	ns
Screat × 2 (*n*)(%)[b]	0 (0)	5 (9)	ns	1 (6)	4 (7)	ns
Failure/relapse (*n*)(%)	2 (13)	10 (18)	ns	2 (12)	10 (18)	ns
Complete remission (*n*)(%)	7 (53)	32 (56)	ns	12 (71)	28 (51)	ns
Partial remission (*n*)(%)	13 (87)	54 (94)	ns	16 (94)	51 (93)	ns
Screat at last follow-up (μmol/l)	86 (71–94)	79 (73–108)	ns	81 (63–111)	79 (72–104)	ns
Proteinuria at last follow-up (g/24 h)	0.5 (0.1–1.4)	0.2 (0.1–0.8)	ns	0.1 (0.1–0.5)	0.3 (0.1–1.1)	ns

IV-S, class IV-segmental; IV-G, class IV-global; CY, patients treated with cyclophosphamide pulses; AZA, patients treated with azathioprine and methylprednisolone pulses; SLEDAI, SLE disease activity index; Screat, serum creatinine; C3, complement C3; C4, complement C4; Screatx2, doubling of serum creatinine. Comparison of ISN/RPS2003 class IV-S and IV-G, and of WHO1995 class IVa/IVb and IVc/IVd. Number and percentage, or median with interquartile range is given.

[a]To convert serum creatinine to mg/dl divide by 88.4.

[b]Primary study end point.

Table 4. Histopathological characteristics of 72 patients with class IV lupus nephritis at study entry

	IV-S $n=15$	IV-G $n=57$	P	IVa or IVb $n=17$	IVc or IVd $n=55$	P
Active lesions						
Endocapillary proliferation	50 (37–77)	80 (67–96)	0.002	86 (68–93)	74 (52–92)	ns
Extracapillary proliferation[a]	21 (5–34)	12 (2–31)	ns	33 (13–41)	10 (2–28)	0.020
Wire loop score ≥25% (*n*) (%)	1 (7)	27 (47)	0.004	8 (47)	20 (36)	ns
Activity index						
Total score	7.7 (5.3–11.7)	9.7 (7.3–12.5)	ns	10.7 (9.2–13.0)	9.3 (7.0–11.0)	0.036
Score ≥10 (*n*) (%)	5 (33)	28 (49)	ns	12 (71)	21 (38)	0.019
Chronic lesions						
Completely sclerosed glomeruli	0 (0–5)	5 (0–9)	ns	0 (0–1)	5 (0–11)	0.002
Synechia	27 (7–38)	16 (6–28)	ns	10 (0–14)	23 (9–38)	0.01
Tubular atrophy score ≥25% (*n*) (%)	1 (7)	6 (11)	ns	0 (0)	7 (13)	ns
Interstitial fibrosis score ≥25% (*n*) (%)	1 (7)	3 (5)	ns	0 (0)	4 (7)	ns
Chronicity index						
Total score	2.3 (1.7–3.3)	2.7 (20–3.7)	ns	2.0 (1.3–2.7)	3.0 (2.0–3.7)	0.000
Score ≥4 (*n*) (%)	1 (7)	8 (14)	ns	0 (0)	9 (16)	ns

Comparison of ISN/RPS2003 class IV-S and IV-G, and comparison of WHO1995 class IVa/IVb and IVc/IVd. Median with interquartile range, or number and percentage are is given.
[a]Fibrous crescents are chronic lesions, while cellular crescents are active lesions. Fibrous crescents were scored rarely.

histopathological characteristics were also evaluated for patients with WHO1995 class IVa or IVb *vs* those classified as class IVc or IVd (data are added to Tables 3 and 4). We argued that class IVa and IVb represent the biopsies with more acute lesions, while IVc and IVd are more chronic forms of LN. Of the 15 patients classified as ISN/RPS2003 class IV-S, only 2 were categorized as WHO1995 class IVa or IVb and 13 as IVc or IVd. For the 57 patients classified as ISN/RPS2003 class IV-G this was 15 and 42, respectively.

Discussion

This is the first time that the interobserver agreement for many predefined histopathological characteristics of LN was studied thoroughly, along with both the WHO1995 and the ISN/RPS2003 classification for LN. Previous studies have hinted on this subject, but either did not incorporate classification data or used the histological characteristics for (prognostic) regression analyses only. The present study was based on 126 renal biopsies of 87 patients with proliferative LN in their initial biopsy, who were included in a randomized controlled trial. We found a wide range of agreement, not only for the diverse histological characteristics but also for the WHO1995 and ISN/RPS2003 classification, of which the latter was assumed to be easily reproducible [5]. A good interobserver agreement was present in only 15% of the 54 scored characteristics, and moderate in 31%. The observation that if more pathologists examine a biopsy, the disagreement will increase, was described in many other pathology studies [11,12], and is therefore not unique for LN. However, since in lupus the variety of histological abnormalities is huge as compared to other diseases, the problem of variable interobserver agreement in lupus will probably be larger. We can only speculate about the reasons for the lack of agreement.

The agreement of the total score of AI and CI appeared to be good to moderate (0.716 and 0.494, respectively). This is in line with observations by others [13].

During the scoring sessions and the consensus meetings, we encountered several difficulties in interpreting the definitions and slides. Although definitions for each item were agreed upon beforehand, the nephropathologists may have interpreted some of these definitions differently during the scoring rounds. Most importantly, however, some of the lesions seem to defy definition, of which we give a number of examples in Figure 1.We were somewhat hampered by the variable quality of the slides, caused by the fact that the biopsies were taken, cut and stained locally. Unfortunately, immunofluorescence slides were not available for this study.

The disagreement on classification was striking: both for the WHO1995 and ISN/RPS2003 classification low ICCs were found (<0.2). Potential causes for these low levels of agreement are: poor quality of the slides, smaller biopsies, biopsies in which around half of the glomeruli were affected, affected glomeruli with both segmental and global lesions, and the presence of the most prominent proliferative lesions in the glomeruli at the edge of the slide. According to the definition of total number of glomeruli, the latter had to be disregarded, thereby hampering a proper classification. No clinical data were given to the pathologists. One could argue that disagreement over a class III or IV LN is less cumbersome than disagreement over class II or IV. Therefore, we categorized the biopsies in proliferative *vs* non-proliferative LN. This did not reveal better ICCs (ICC for WHO1995 was 0.312 and for ISN/RPS2003 was 0.251). So the discrepancies in classification cannot be explained by disagreement

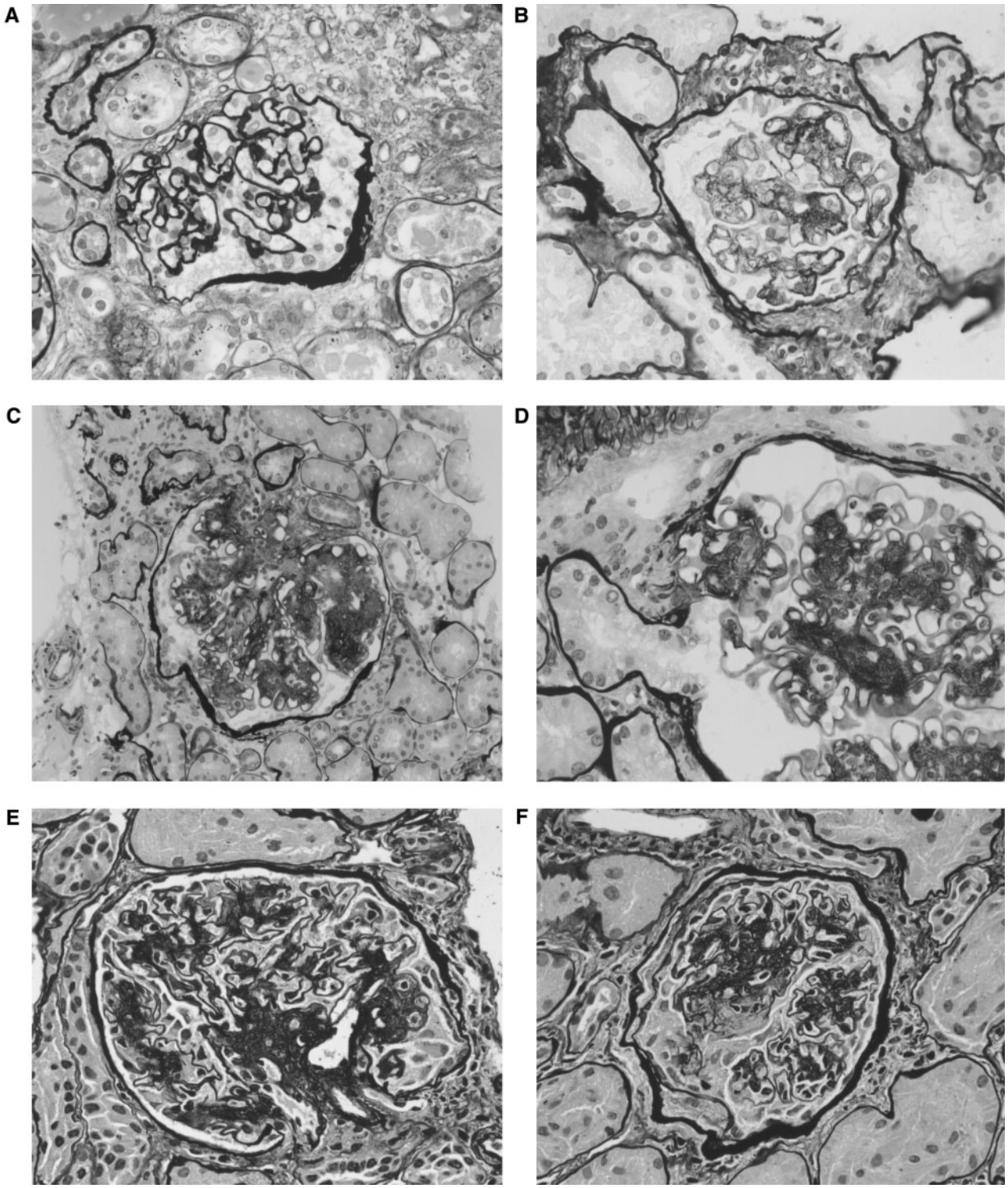

Fig.1. Several examples of histomorphological characteristics, which might explain why a low interobserver agreement was found. Furthermore, since the ISN/RPS2003 classification makes an explicit distinction between segmental and global lesions, we would like to show you how difficult it can sometimes be to make this distinction. Considerations/answers are given in parentheses following the questions. (**A**) Is this a normal glomerulus? (This is clearly not a normal glomerulus because it shows signs of ischemia, namely wrinkling of GBM and irregular Bowman's capsule. However, it does not contain lupus-related lesions. In counting glomeruli with and without lupus-related lesions, for instance, to make a decision on a class III or class IV, this glomerulus would be on the 'normal' side, and was categorized as such on the scoring form.) (**B**) Glomerulus with mesangial changes, but is there endocapillary proliferation as well? (Next to mesangial changes this glomerulus shows chronic lesions, namely doubling of GBM and ischemia. However, in a number of capillary loops, endothelial cells are prominently present. However, leucocytes are absent and obliteration of capillary loops due to proliferation hardly occurs. Three out of five pathologists scored this as segmental endocapillary proliferation. However, the other two scored it as global endocapillary proliferation. It is certainly a case of A/C.) (**C**) Glomerulus with extensive mesangial proliferation, but is there a component of endocapillary proliferation

regarding focal or diffuse LN. Finally, for a correct classification, a minimum number of ten glomeruli is required [4]. In several cases, our classification was made on less than ten glomeruli. However, this did not explain the low interobserver agreement, since ICCs did not change if the data were analysed for biopsies with ten or more glomeruli only. Strikingly, we found that during the consensus meetings, consensus was often reached without difficulties.

Reports on the interobserver agreement of the WHO1995 or ISN/RPS2003 classification are scarce [14,15]. In most studies on LN only one nephropathologist classified the biopsies [16,17], or two nephropathologists classified the biopsies in consensus [6,18,19]. An exception is a report on the ISN/RPS2003 classification which described 60 Japanese patients with LN [14]. The consensus between two pathologists for the ISN/RPS2003 and WHO1995 classification was 98 and 83%, respectively. Another report showed that the consensus between two pathologists who studied 420 biopsies, was better for the ISN/RPS2003 classification than for the WHO1995 classification (96% *vs* 87%) [20]. Recently, Furness *et al*. reported a study on interobserver variability throughout the UK [15]. Twenty renal biopsies, accompanied with clinical information and immunofluorescence results, were sent to (untrained) pathologists twice: the first time the pathologist was asked to classify the biopsies according to the WHO1995 classification, the second according to the ISN/RPS2003 classification. Calculated kappa values were 0.44 and 0.53 for the WHO1995 and ISN/RPS2003 classification, respectively. Disease activity (acute *vs* chronic *vs* acute and chronic) was poorly agreed upon (κ 0.33). They concluded that the ISN/RPS2003 classification is better reproducible. In our study, the pathologists were unaware of clinical data.

The differences between class IV-S and IV-G may become clinically important and lead to different therapeutical guidelines. Our patients with IV-G had more active disease and worse clinical parameters at study entry. The outcome however was not different, and none of the patients who were classified as IV-S doubled their serum creatinine. One of the reasons that we did not find differences in outcome could be due to the fact that doubling of serum creatinine did not occur often in our study ($n=5$), after a reasonable follow-up. Like others [14,21,22], we could not confirm the findings by Najafi *et al*. [6] that IV-S had a worse prognosis than IV-G. All these studies [14,21,22] also found worse clinical and histopathological characteristics in the IV-G patients. The chronicity index in our population was lower than that in the original study [6], and higher than that found in the French study [22]. It is important to realize that the initial report described an American population, while most of the recent studies, including ours, were carried out on either European or Asian patients. The comparison of patients with WHO1995 class IVa or IVb with those classified as class IVc or IVd revealed that IV-S and IV-G present as different clinical entities, while IVa/b and IVc/d are histopathologically different. Since for the patients with class III LN an impaired renal function was obligatory for inclusion in our trial, a comparison between the patients with class IV-S and IV-G with those with class III ($n=11$) is not meaningful.

In summary, although standardizing the assessment of renal biopsies in LN is needed both for clinical and research purposes, and although the new ISN/RPS2003 classification seems to be of help, our study shows that the detailed evaluation by specialized nephropathologists shows variable degrees of agreement with regard to various histopathological characteristics. However, the detailed scoring system we used is too laborious and unpractical for daily clinical use.

Our results do not weaken the importance of the classification for LN, but they raise the question whether a further subclassification is meaningful. Indeed, we did not find differences in the outcome between the patients with class IV-S or IV-G lesions. We believe that, until proven clinically useful, subclassifying diffuse proliferative LN into segmental or global should be used for research purposes only. Additional research has to reveal whether or not these groups are (clinically) different and should be treated differently.

Acknowledgements. We would like to thank all patients and the members of the Dutch Working Party on SLE who were willing to participate in the first Dutch Lupus Nephritis Study. This study was possible because of the skilful coordination by Dr G. Ligtenberg (Department of Nephrology, and Department of Rheumatology & Clinical Immunology, University Medical Center Utrecht, Utrecht, The Netherlands). We thank the pathologists who supplied the renal biopsies. Furthermore, we have appreciated the help of Drs J. J. Weening (Department of Pathology, Academic Medical Center, Amsterdam, The Netherlands), J. A. Bruijn (Department of

as well? (The mesangial proliferation is dominant and most of the capillary loops surround the mesangial areas like a corona, without apparent endocapillary proliferation. However, there seems to be endocapillary proliferation at 7 o'clock, and also at 9 o'clock, to be scored as segmental.) (**D**) Glomerulus with a segmental proliferative lesion, but is it active or chronic? (The mesangial changes in the background of the proliferative lesions are disregarded in making this decision. The proliferative lesion next to the proximal tube has an element of chronicity, namely its adhesion to the capsule, and of activity, namely its active endocapillary proliferation. Two pathologists scored it as a chronic lesion, letting the adhesion prevail on the decision. The others chose for A/C.) (**E**) Glomerulus with mesangial sclerosis, but is it global or segmental? According to the definition, global mesangial sclerosis involves more than 50% of the glomerular tuft, segmental mesangial sclerosis less than 50%. (Mesangial sclerosis is most prominently present in the center and on the right side of the glomerulus. To a lesser extent, it is present on the left side and upper side as well. Does it reach an area of more than 50%? All pathologists decided it did.) (**F**) Glomerulus with minor inflammation, enough for a positive score? According to the definition, at least five inflammatory cells should be present. (There seem to be five inflammatory cells present in capillary loops, so the answer is yes, and all pathologists agreed on this. However, some inflammatory cells are difficult to recognize, and could easily be mistaken for endothelial cells. There are also cells with apoptotic changes, of which it is uncertain whether they are inflammatory cells.)

Pathology, Leiden University Medical Center, Leiden, The Netherlands) and K. J. Assmann (Department of Pathology, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands), who participated in the initial development of the scoring form. Finally, we would like to thank Dr G. S. Hill (Department of Pathology, Hôpital Europeen Georges Pompidou, Paris, France) for his contribution to the scoring form.
This study was supported by grants from the Dutch Kidney Foundation (C02.2023 and C94.1363) and the Dutch Arthritis Association (No.735). C.G. is a recipient of a Netherlands Organization for Scientific Research Fellowship for Clinical Investigators (NWO No. 920-03-115).

Conflict of interest statement. None declared.

References

1. Wallace DJ, Hahn BH. *Dubois' Lupus Erythematosus*. Williams&Wilkins, Baltimore, MD: 1997
2. Nossent JC, Bronsveld W, Swaak AJ. Systemic lupus erythematosus. III. Observations on clinical renal involvement and follow up of renal function: Dutch experience with 110 patients studied prospectively. *Ann Rheum Dis* 1989; 48: 810–816
3. Churg J, Bernstein J, Glassock RJ. *Lupus nephritis. Renal Disease: Classification and Atlas of Glomerular diseases*. Igaku-Shoin, New York: 1995; 151–180
4. 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
5. Chan TM. Histological reclassification of lupus nephritis. *Curr Opin Nephrol Hypertens* 2005; 14: 561–566
6. Najafi CC, Korbet SM, Lewis EJ, Schwartz MM, Reichlin M, Evans J. Significance of histologic patterns of glomerular injury upon long-term prognosis in severe lupus glomerulonephritis. *Kidney Int* 2001; 59: 2156–2163
7. Grootscholten C, Ligtenberg G, Hagen EC *et al*. Azathioprine/methylprednisolone versus cyclophosphamide in proliferative lupus nephritis. A randomized controlled trial. *Kidney Int* 2006; 70: 732–742
8. Austin HA III, Muenz LR, Joyce KM, Antonovych TT, Balow JE. Diffuse proliferative lupus nephritis: identification of specific pathologic features affecting renal outcome. *Kidney Int* 1984; 25: 689–695
9. Grootscholten C, Bajema IM, Florquin S *et al*. Treatment with cyclophosphamide delays the progression of chronic lesions more effectively than does treatment with azathioprine plus methylprednisolone in patients with proliferative lupus nephritis. *Arthritis Rheum* 2007; 56: 924–937
10. Fleiss JL. The measurement of interrater agreement. In: Fleiss JL, ed. *Statistical Methods for Rates and Proportions*. John Wiley & Sons, New York: 1981; 212–236
11. Bajema IM, Hagen EC, Hansen BE *et al*. The renal histopathology in systemic vasculitis: an international survey study of inter- and intra-observer agreement. *Nephrol Dial Transplant* 1996; 11: 1989–1995
12. Furness PN, Taub N. International variation in the interpretation of renal transplant biopsies: report of the CERTPAP Project. *Kidney Int* 2001; 60: 1998–2012
13. Wernick RM, Smith DL, Houghton DC *et al*. Reliability of histologic scoring for lupus nephritis: a community-based evaluation. *Ann Intern Med* 1993; 119: 805–811
14. 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
15. 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
16. Bajaj S, Albert L, Gladman DD, Urowitz MB, Hallett DC, Ritchie S. Serial renal biopsy in systemic lupus erythematosus. *J Rheumatol* 2000; 27: 2822–2826
17. Hill GS, Delahousse M, Nochy D *et al*. Outcome of relapse in lupus nephritis: roles of reversal of renal fibrosis and response of inflammation to therapy. *Kidney Int* 2002; 61: 2176–2186
18. Austin HA III, Boumpas DT, Vaughan EM, Balow JE. High-risk features of lupus nephritis: importance of race and clinical and histological factors in 166 patients. *Nephrol Dial Transplant* 1995; 10: 1620–1628
19. Moroni G, Pasquali S, Quaglini S *et al*. Clinical and prognostic value of serial renal biopsies in lupus nephritis. *Am J Kidney Dis* 1999; 34: 530–539
20. Li H, Li X-W. The clinical usefulness of new ISN/RPS2003 classification in 420 biopsy-proven lupus nephritis. *J Am Soc Nephrol* 2005; 16: 525Ref Type: Abstract
21. Mittal B, Hurwitz S, Rennke H, Singh AK. New subcategories of class IV lupus nephritis: are there clinical, histologic, and outcome differences? *Am J Kidney Dis* 2004; 44: 1050–1059
22. Hill GS, Delahousse M, Nochy D, Bariety J. Class IV-S versus class IV-G lupus nephritis: Clinical and morphologic differences suggesting different pathogenesis. *Kidney Int* 2005; 68: 2288–2297

Received for publication: 24.6.07
Accepted in revised form: 20.7.07