Diabetic nephropathy : from histological findings to clinical features
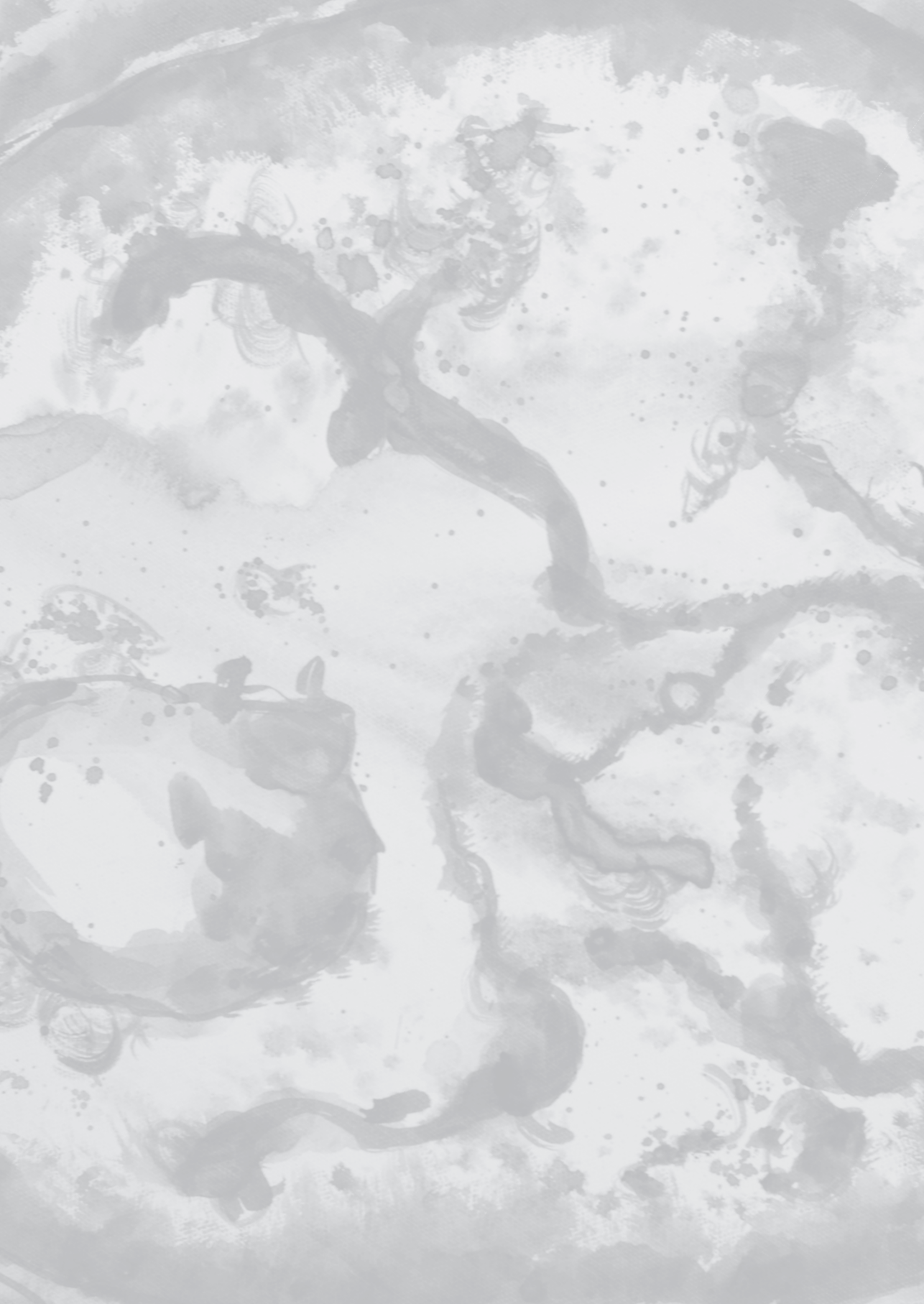
Klessens, C.Q.F.

Citation

Klessens, C. Q. F. (2017, November 22). *Diabetic nephropathy : from histological findings to clinical features*. Retrieved from https://hdl.handle.net/1887/55808

Version:	Not Applicable (or Unknown)
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/55808

Note: To cite this publication please use the final published version (if applicable).

Pathologic classification of diabetic nephropathy; reproducibility and validation

Elisabeth J.J. Valk
Celine Q.F. Klessens
Olaf M. Dekkers
Ron Wolterbeek
Kerstin Amman
Adam H. Cohen
Terry H. Cook
Cinthia B. Drachenberg
Franco Ferrario
Agnes B. Fogo
Kensuke Joh
Laure-Helene Noël
Jai Radhakrishnan
Surya V. Seshan
Jan A. Bruijn
Mark Haas
Ingeborg M. Bajema
Antien L. Mooyaart

4

ABSTRACT

In 2010, a pathological classification for diabetic nephropathy (DN) was developed. Most validation studies of this classification showed a significant association with renal outcome, but either compared different combinations of classes with outcome or had insufficient power. An adequate reproducibility study of this classification system is missing. Therefore, we performed a reproducibility study together with a meta-analysis study for all validation studies to better estimate the prognostic role of the different DN classes.

In order to asses agreement for reproducibility, a DVD with 13 digitized biopsies of DN patients was sent to all members of the Renal Pathology Society. The interobserver agreement was determined by intraclass correlation values (ICC). Additionally, data was extracted from the validation studies and a meta-analysis for the different DN classes was performed.

The ICC for DN class was 0.74. Additional parameters had the following ICCs: IFTA: 0.72; interstitial inflammation: 0.47; arteriolar hyalinosis: 0.41 and arteriosclerosis: 0.43. ICCs of all pathologists did not differ substantially from ICCs of pathologists with a high level of expertise. The meta-analysis showed an increased risk for a poor renal outcome for each of the classes IIb, III and IV compared to class IIa ($p<0.0001$).

This study shows that the DN classification is suitable for clinical practice. Our meta-analysis data show a good relation with renal prognosis over the classes of DN as currently defined. Furthermore, the DN classification has relatively good reproducibility although improvements could be made by fine-tuning definitions.

INTRODUCTION

Pathological classification systems exist for several renal diseases such as lupus nephritis [1], IgA nephropathy [2] and ANCA-associated glomerulonephritis [3]. In 2010, a classification system was developed for diabetic nephropathy (DN) [4]. This classification is primarily based on glomerular damage whereas interstitial and vascular lesions are scored separately. The classification system can be used for DN in both type 1 and type 2 diabetes mellitus patients.

In general, classification systems are created to provide better communication between pathologists and clinicians, and the possibility to link diagnostic information together with prognostic indication in order to guide therapeutic decision making. By means of subsequent studies focusing on the clinical validation and interobserver agreement of classification systems, the systems are continuously modified and improved.

For the pathological classification of DN, several validation studies of the DN classes showed an association with renal outcome. However, all these studies combined classes, and moreover, combined them differently to have adequate power to find an association with outcome [5-9]. The reproducibility of the DN classification was investigated in the original article and one validation study, but these reproducibility studies were relatively small [4, 8].

Therefore, the aim of the present study was to perform an adequate reproducibility study by means of a survey of members of the Renal Pathology Society (RPS). Based on the data obtained, we here provide suggestions for clarifications of some of the definitions. Furthermore, we performed a meta-analysis study for all validation studies published to date, to adequately estimate the prognostic role of the different DN classes.

METHODS

Reproducibility study; Case selection and survey

For this study 13 biopsies of DN patients were selected from the archives at Leiden University Medical Center. The renal damage within these cases was exclusively attributable to DN. The representative cases contained the full spectrum of lesions which can be found in DN. For each case high quality Periodic Acid-Schiff (PAS) and silver stains were provided. The biopsies were handled, coded and anonymized according to the Dutch National Ethical guidelines.

In order to assess interobserver agreement, slides were digitalized and saved as an HTML-document on a DVD. This DVD was sent to all 360 members of the RPS with an invitation to participate in the project. Additionally, an Excel response sheet with a flowchart of the classification, instructions, and the original article of the classification were provided on the DVD [4]. The participants were asked to add remarks when difficulties occurred while scoring. The participating pathologists were requested to assess their level of expertise by which they were divided into the following categories: all pathologists; the subgroup of pathologists with a high level of expertise; the subgroup of pathologists with low or moderate expertise and those who did not mention their level of expertise by self-assessment.

Pathological classification system of DN

All cases were categorized based on the pathological classification of DN as previously described. For the exact details of the DN classification, we refer to the original paper [4]. In brief, the classification encompasses 4 classes: class I, glomerular basement thickening by electron microscopy without specific light microscopic changes; class IIa, mild mesangial expansion (mild mesangial expansion in >25% of the observed mesangium); class IIb, severe mesangial expansion (severe mesangial expansion in >25% of the observed mesangium); class III, at least one lesion with nodular sclerosis; class IV global glomerulosclerosis in > 50% of glomeruli.

Interstitial fibrosis and tubular atrophy (IFTA) and interstitial inflammation are scored on a semiquantitative scale. Arteriolar hyalinosis is scored 0 when it was absent; 1, if at least one arteriole with hyalinosis is present and 2, if more than one arteriole with hyalinosis is observed in the entire biopsy. Arteriosclerosis is scored as follows: 0 for no intimal thickening, 1 for intimal thickening less than the thickness of the media, and 2 for intimal thickening more than the thickness of the media.

Eligibility criteria, data selection and extraction of validation studies

To provide an overview of clinical outcome studies in relation to the pathological classification of DN, all articles citing the original manuscript were collected via a search on Web of Science and Google Scholar. To ensure maximum sensitivity, no limits or filters were used in the searches. Language restrictions were not included in the initial search. This search was performed by a trained librarian in April 2016. Two observers (CK and LV) independently reviewed all studies to include all validation studies of the pathological classification of DN, which associated DN class with renal outcome. Furthermore, an expert of the field (IB) was consulted to ensure that all validation studies investigating DN class were included. Included were studies investigating type 1 and 2 diabetes correlating DN class with renal outcome. Renal outcome was defined as end stage renal disease (ESRD) or doubling of se-

rum creatinine. Studies were excluded which only investigated interstitial lesions or which used the same cohort in relation to different clinical outcome parameters. In the latter case, we included the first published validation study in our meta-analysis. Data were extracted if possible. Preferably hazard ratios were extracted with confidence intervals, otherwise these data were extrapolated from Kaplan-Meier curves or data on absolute risks.

Statistical analysis

For all the parameters in this study a reliability analysis was conducted by calculating the intraclass correlation coefficient (ICC) (0=no agreement, 1=perfect agreement). In this calculation, the given answers were compared between participants rather than comparing the answers with a 'gold standard'. ICCs were calculated using a mixed model to estimate the variance components of the ICC. An ICC of >0.75 was considered to show excellent reproducibility, ICC of 0.4 to 0.75 to indicate fair to good reproducibility, and ICC of < 0.4 to indicate poor reproducibility [10]. These analyses were performed using SPSS statistics 20.0 (IBM, Armonk, NY). For the meta-analysis of the validation study a generic-invariance method was used in a random effects model and analyses were performed in ReviewManager (RevMan) version 5.3. Preferably hazard ratios from the individual studies were used with class IIa as a reference group. If these were not available, relative risks and standard errors were calculated with the available data. Heterogeneity within the studies was estimated by the I^2, which is the percentage of the total variation across studies due to heterogeneity rather than chance. An I^2 of 25%, 50% or 75% was considered low, moderate or high, respectively.

RESULTS

Reproducibility study

A total of 13 biopsies with lesions attributable to DN were scored by 77 pathologists from 28 different countries, of which 38 (49%) had a self-assessed high level of expertise, 19 (25%) had a moderate level, 3 (4%) had a low level and 21 (22%) did not mention their level of expertise. The response rate of the reproducibility study was 21.4% (77/360).

Table 1. Intraclass correlation (ICC) of pathologic classification of DN

ICC score	All pathologists	Pathologists with high expertise level
DN class	0.74	0.76
IFTA	0.72	0.73
Interstitial inflammation	0.47	0.57
Arteriolar Hyalinosis	0.41	0.43
Arteriosclerosis	0.43	0.44

ICCs of all pathologists did not differ substantially from ICCs of pathologists with a high level of expertise

Table 1 gives an overview of intraclass correlations (ICC) of the lesions scored by all participating pathologists and the ICC of a subgroup of pathologists with a high level of expertise. ICCs of all pathologists did not differ substantially from ICCs of pathologists with a high level of expertise The ICC score for glomerular lesions (i.e. DN class) amongst all pathologists was 0.74; amongst pathologists with high expertise it was 0.76. The best concordance was found in class III and IV; the least concordance was found in class I and II, and in the subdivision of class II.

IFTA had an ICC score of 0.72. Severe IFTA cases had more concordance compared to cases with mild IFTA involvement. Regarding the vascular lesions, arteriolar hyalinosis had an overall reproducibility of 0.41. Arteriosclerosis had an overall interobserver agreement of 0.43. The participating members of the RPS could provide separate remarks in addition to the scoring system. Table 2 provides an overview of certain remarks, followed by our recommendations to clarify these definitions.

Meta-analysis on validation studies

The initial search found 258 studies, of which 240 studies were excluded because these were case reports, reviews, or studies which used the classification system in an experimental setting, but were not validation studies. Finally, 12 studies were regarded as possible validation studies, however, some only investigated IFTA with renal outcome or the same cohort was used in more than one study. Therefore, in the end, 4 validation studies were included in the meta-analysis (Figure 1). The basic characteristics of these studies are provided in Table 3. In the study of Mise *et al.* and Okada *et al.* data including the hazard ratios with confidence interval (of which the standard error could be calculated) were available. In the study by Okada *et al.* no data for class I were available. In the study by Oh *et al.* the described absolute risks were used to calculate an relative risk and standard error. In the study by An *et al.* there was a hazard ratio and confidence interval available for all classes together. The hazard ratio was extrapolated for all independent classes from this hazard ratio and the standard error was calculated from the relative risk which was extrapolated from the Kaplan-Meier curves.

Table 2. Recommendations survey for DN classification

		Remarks	Recommendations
Glomerular lesions	*DN class*	Dividing class I and II: definition and cut-off points criticized as difficult	Define more straightforward definitions for mesangial alterations and examine these in future validation studies.
		Class III: only one nodule in cases with overall mild mesangial expansion	The formation of nodular sclerosis may be a specific trait of some patients with DN, who are not yet more distinctly defined Therefore, recognition of one nodule seems appropriate to designate a specific class of DN. Results from the meta-analysis indicate specific outcome for this class.
Interstitial lesions	*IFTA*	IFTA is a good predictor for renal outcome, but can be the result of other renal diseases	Take the severity of IFTA into account and specifically note in all biopsy reports
	Interstitial inflammation	The relatively low ICC of interstitial inflammation	Clarification of this definition: only score inflammation in areas without IFTA. Determine the relative effects of interstitial inflammation in non-scarred areas versus total interstitial inflammation on the prognosis of DN in further studies.
Vascular lesions	*Arteriolar hyalinosis*	Unclear in which vessel type arteriolar hyalinosis needs to be scored	Just identify and mention hyalinosis if present; most of the hyalinosis will occur in arterioles.
	Arteriosclerosis	What is the definition of a large vessel	Eliminate the vessel size from the definition, and to focus only on the presence of intimal thickening/fibrosis in vessels which are larger than arterioles

Based on the remarks obtained from the reproducibility study, recommendations of each parameter were proposed to improve the use of the DN classification

4

Figure 2 shows the results of our meta-analyses of the validation studies. The validation studies resulted in a pooled hazard ratio of class I versus IIa of 0.49 (95% C.I. 0.13-1.90, p=0.30). The pooled hazard ratio of class IIa versus class IIb was 2.96 (95% C.I. 1.82-6.05, p<0.00001), showing an increased risk for developing a poor renal outcome in patients with class IIb compared to IIa. For class IIa versus class III a significant difference also was seen (p<0.00001) with a pooled hazard ratio of 5.26 (95% C.I. 2.75-10.04, p<0.00001), showing that patients with class III have an increased risk of a poor renal outcome compared to class IIa. Finally, class IV versus IIa showed a poorer renal prognosis for class IV, hazard ratio 11.23 (95% C.I. 4.56-27.68, p<0.00001). There is low to moderate heterogeneity between the groups (class IIa vs I, I^2= 0%, class IIa vs IIb, I^2= 0%, class IIa vs III, I^2= 49% and class IIa vs IV, I^2= 67%).

Figure 1. Flowchart illustrating how the validation studies of the DN classification were selected for the meta-analysis
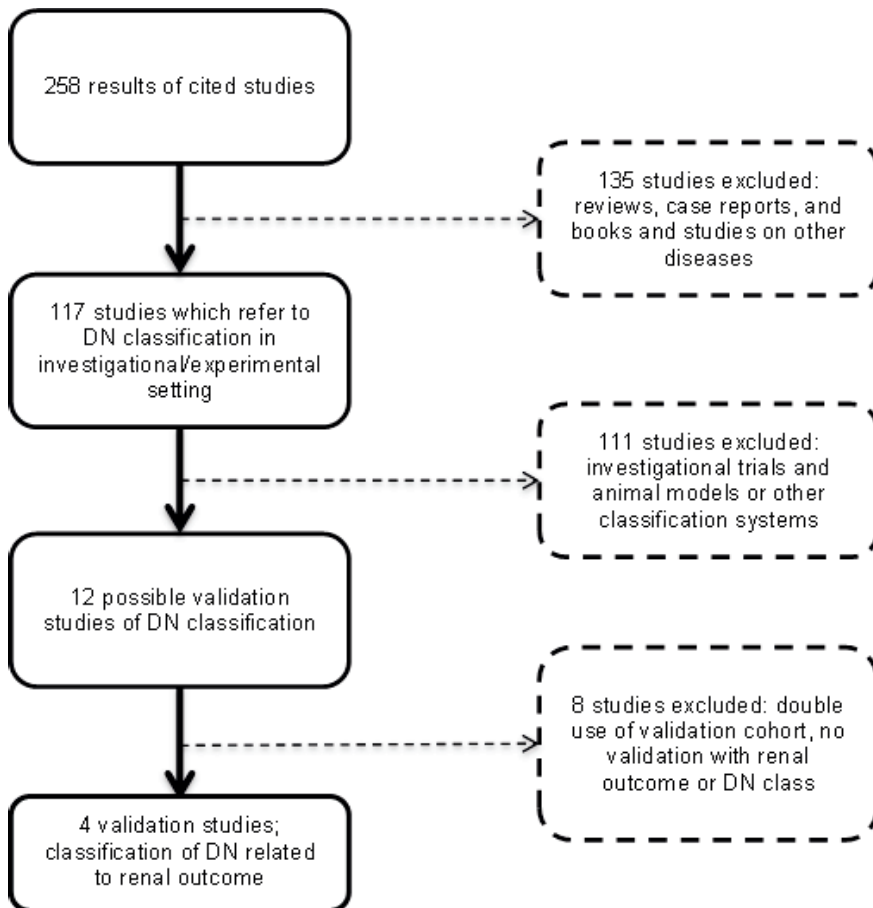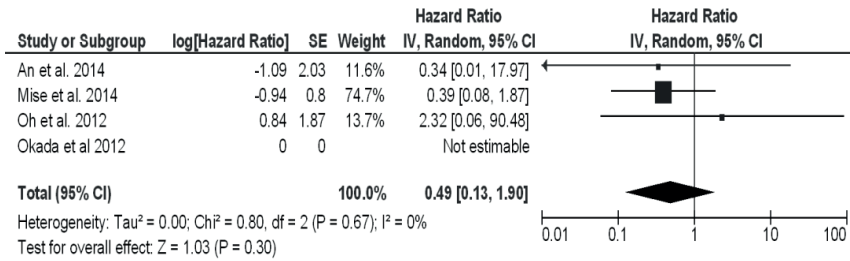
Table 3. Overview of validation studies using the histopathological classification of DN
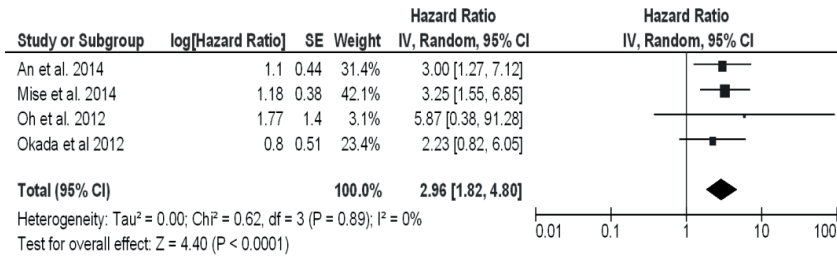
Study	Diabetes type	Number of patients	Ethnicity	Patient characteristics	Follow up	Renal outcome
Okada *et al.* 2011	Type 2 diabetes	N= 69	Japanese	Patients with overt proteinuria and biopsy-confirmed diabetic nephropathy with mesangial expansion	Mean and median follow-up duration was 59 +/-41 and 52 months (range 6 – 180 months)	Chronic dialysis or doubling of serum creatinine
Oh *et al.* 2012	Type 2 diabetes	N= 126	Korean	50 patients with pure DN, 65 with non diabetic renal disease and 11 mixed. Only 50 pure DN patients scored following the DN classification	Follow-up during 69.2 +/-35.2 (0.4–137.6) months after renal biopsy to detect ESRD	End-stage renal disease
Mise *et al.* 2014	Diabetes mellitus	N= 205	Japanese	Patients with renal biopsy and inclusion study criteria (eGFR>10mL/min/1.73m², biopsy>10 glomeruli)	The mean follow-up period was 62.9 ± 68.3 months	Renal death defined as dialysis by end-stage renal disease
An *et al.* 2014	Type 2 diabetes	N= 396	Chinese	Patients with biopsy proven DN	At least one year follow up	Renal outcome defined as progression to end-stage renal disease or doubling of serum creatinine

4

Figure 2. Meta-analysis of DN classes of the performed validation studies at 80 month follow-up
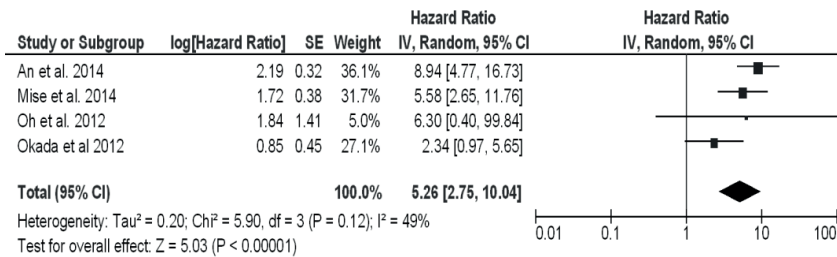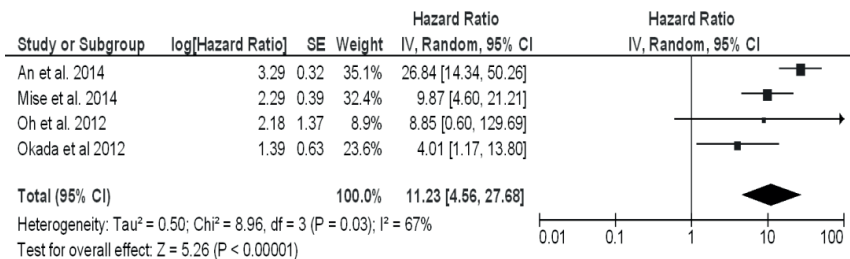
IIa v I

Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Random, 95% CI
An et al. 2014	-1.09	2.03	11.6%	0.34 [0.01, 17.97]
Mise et al. 2014	-0.94	0.8	74.7%	0.39 [0.08, 1.87]
Oh et al. 2012	0.84	1.87	13.7%	2.32 [0.06, 90.48]
Okada et al 2012	0	0		Not estimable
Total (95% CI)			**100.0%**	**0.49 [0.13, 1.90]**

Heterogeneity: Tau² = 0.00; Chi² = 0.80, df = 2 (P = 0.67); I² = 0%
Test for overall effect: Z = 1.03 (P = 0.30)

IIa vs IIb

Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Random, 95% CI
An et al. 2014	1.1	0.44	31.4%	3.00 [1.27, 7.12]
Mise et al. 2014	1.18	0.38	42.1%	3.25 [1.55, 6.85]
Oh et al. 2012	1.77	1.4	3.1%	5.87 [0.38, 91.28]
Okada et al 2012	0.8	0.51	23.4%	2.23 [0.82, 6.05]
Total (95% CI)			**100.0%**	**2.96 [1.82, 4.80]**

Heterogeneity: Tau² = 0.00; Chi² = 0.62, df = 3 (P = 0.89); I² = 0%
Test for overall effect: Z = 4.40 (P < 0.0001)

IIa vs III

Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Random, 95% CI
An et al. 2014	2.19	0.32	36.1%	8.94 [4.77, 16.73]
Mise et al. 2014	1.72	0.38	31.7%	5.58 [2.65, 11.76]
Oh et al. 2012	1.84	1.41	5.0%	6.30 [0.40, 99.84]
Okada et al 2012	0.85	0.45	27.1%	2.34 [0.97, 5.65]
Total (95% CI)			**100.0%**	**5.26 [2.75, 10.04]**

Heterogeneity: Tau² = 0.20; Chi² = 5.90, df = 3 (P = 0.12); I² = 49%
Test for overall effect: Z = 5.03 (P < 0.00001)

IIa vs IV

Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Random, 95% CI
An et al. 2014	3.29	0.32	35.1%	26.84 [14.34, 50.26]
Mise et al. 2014	2.29	0.39	32.4%	9.87 [4.60, 21.21]
Oh et al. 2012	2.18	1.37	8.9%	8.85 [0.60, 129.69]
Okada et al 2012	1.39	0.63	23.6%	4.01 [1.17, 13.80]
Total (95% CI)			**100.0%**	**11.23 [4.56, 27.68]**

Heterogeneity: Tau² = 0.50; Chi² = 8.96, df = 3 (P = 0.03); I² = 67%
Test for overall effect: Z = 5.26 (P < 0.00001)

DISCUSSION

In 2010, we launched the histopathological classification for diabetic nephropathy, which since then has been used in multiple research and diagnostic settings. In general practice, nephropathologists worldwide make use of the classification, but fine-tuning of some of the definitions of lesions may be appropriate. Therefore, a survey was launched through the RPS to obtain insight into interobserver agreement amongst nephropathologists worldwide, and issues related to day-to-day practical issues. With regards to definitions, we here summarize the comments from RPS members who joined in our survey. Whereas reproducibility was proven to be sufficient for classes III and IV, disagreement among observers was noticed for class I, IIa and IIb. Although many studies citing the classification are present in the literature, a careful investigation showed that actually only a limited number of validation studies had been conducted, i.e. those with a specific aim to validate the classes of the DN classification. All these studies came from Asia. We also performed a meta-analysis of these studies, showing that the DN classification has a correlation with renal outcome for most classes using class IIa as a reference, which underscores the clinical usefulness of the classification system. For class I versus IIa there was no significant difference found in this meta-analysis. This is likely explained by the fact that the available studies were underpowered for this comparison as only a small number of patients were investigated with only few events (especially in class I).

Because the meta-analysis showed a good association with renal outcome for most classes, it seems appropriate to maintain the subdivision as previously described. However, more straightforward definitions for mesangial alterations to distinguish between class I, IIa and IIb are called for, but need to be examined in future validation studies using modifications of the original definitions that would be tested for both interobserver agreement and correlation with clinical outcomes. For the moment, no clear-cut suggestions came out of the survey for an intermediate solution.

With respect to classes III and IV, there was good interobserver agreement in the recognition of these classes and comments from participants were few. There was a concern from some participants whether the presence of one nodule was sufficient to classify a sample as class III DN, especially in cases with overall mild mesangial expansion. We postulate that the formation of nodular sclerosis may be a specific trait of some patients with DN, who are not yet more distinctly defined [11]. Therefore, recognition of one nodule still seems appropriate to designate a specific class of DN (i.e. class III), and results from the meta-analysis indicate a specific outcome for this class by its current definition.

Our reproducibility study showed a high ICC for IFTA, similar to findings in the original Oxford IgA nephropathy study [2]. A common remark from survey participants was that IFTA could have been the result of other renal diseases. The point about IFTA being a good predictor for renal outcome despite its nonspecific appearance in virtually all renal diseases has been frequently raised. IFTA is not a primary parameter of the DN classification, but several validation studies on IFTA showed that IFTA has impact on renal prognosis in diabetic nephropathy [5, 7-9, 12]. Because of the prognostic value of IFTA it is certainly useful to take the severity of IFTA into account during evaluation of the biopsy, and the amount of IFTA should be specifically noted in all renal biopsy reports in cases of DN. The ICC of interstitial inflammation was remarkably low. This could be the result of lack of clarity in the definition of whether or not to score inflammation in areas with or without IFTA. In concordance with other guidelines for scoring inflammation in renal diseases, it would seem appropriate also in DN only to score inflammation in areas without IFTA. Future studies need to clarify the definition of interstitial inflammation and should determine the relative effects of interstitial inflammation in non-scarred areas versus total interstitial inflammation on the prognosis of diabetic nephropathy.

According to the original manuscript of the DN classification, arteriosclerosis needs to be scored in large vessels. However, the definition of a large vessel was perceived as unclear. A straightforward and simple solution would be to eliminate the vessel size from the definition, and to focus only on the presence of intimal thickening/fibrosis in vessels which are larger than arterioles. In addition, the vessel type and/or size in which arteriolar hyalinosis needs to be scored was also not evident. A straightforward solution here would be just to identify and mention hyalinosis if present – given that the lesion is relatively easy to recognize in a PAS-staining – and most of the hyalinosis lesions will occur in arterioles.

In the present study we reflect on problematic issues of the DN classification by using a two-way approach, namely through a survey of renal pathologists from the RPS and by reviewing the literature by means of a meta-analysis of validation studies so far performed. Each of these routes has its own limitations. The survey approach may have suffered from a bias in response due to factors outside our control, because RPS members could become involved in this part of the study by their own initiative. Nevertheless, participants in this study on diabetic nephropathy appeared to reflect the RPS membership relatively well with regards to the participants' distribution among different countries and with respect to the range of experience among our participants. During the evaluation of the validation studies included in our meta-analysis, we observed some limitations of these studies that merit discussion. The performed validation studies had different study designs and all glomerular classes were not included in

all studies. For example, Okada *et al.* did not include class I in their study [7]. Some of the studies, especially by Oh *et al.* and Okada *et al.* were very small. Furthermore, all studies were performed in Asian populations, and therefore it could be debated if these data can be fully extrapolated to other populations.

This study shows that the pathologic classification of DN has relatively good reproducibility but improvements can be made by fine-tuning definitions. Our meta-analysis data showed a good relation with renal prognosis over the classes of DN as currently defined. On the basis of results from a two-way approach into reproducibility and prognostic value of the classification, we have listed the current issues with recommendations here. In an international workgroup on DN we are currently working on the modifications of the DN classification.

ACKNOWLEDGEMENTS

REFERENCES

1. Weening JJ, D'Agati VD, Schwartz MM, et al. The classification of glomerulonephritis in systemic lupus erythematosus revisited. *Kidney Int*, 2004; 65: 521-30

2. Working Group of the International IgA Nephropathy Network and the Renal Pathology Society. The Oxford classification of IgA nephropathy: pathology definitions, correlations, and reproducibility. *Kidney Int*, 2009; 76: 546-56

3. Berden AE, Ferrario F, Hagen EC, et al. Histopathologic classification of ANCA-associated glomerulonephritis. *J Am Soc Nephrol*, 2010; 21: 1628-36

4. Tervaert TW, Mooyaart AL, Amann K, et al. Pathologic classification of diabetic nephropathy. *J Am Soc Nephrol*, 2010; 21: 556-63

5. Mise K, Hoshino J, Ubara Y, et al. Renal prognosis a long time after renal biopsy on patients with diabetic nephropathy. *Nephrol Dial Transplant*, 2014; 29: 109-18

6. Oh SW, Kim S, Na KY, et al. Clinical implications of pathologic diagnosis and classification for diabetic nephropathy. *Diabetes Res Clin Pract*, 2012; 97: 418-24

7. Okada T, Nagao T, Matsumoto H, et al. Histological predictors for renal prognosis in diabetic nephropathy in diabetes mellitus type 2 patients with overt proteinuria. *Nephrology (Carlton)*, 2012; 17: 68-75

8. An Y, Xu F, Le W, et al. Renal histologic changes and the outcome in patients with diabetic nephropathy. *Nephrol Dial Transplant*, 2015; 30: 257-66

9. Zhu X, Xiong X, Yuan S, et al. Validation of the interstitial fibrosis and tubular atrophy on the new pathological classification in patients with diabetic nephropathy: A single-center study in China. *J Diabetes Complications*, 2016; 30: 537-41

10. Fleiss JL and Cohen J. Equivalence of Weighted Kappa and Intraclass Correlation Coefficient as Measures of Reliability. *Educational and Psychological Measurement*, 1973; 33: 613-619

11. Schwartz MM, Lewis EJ, Leonard-Martin T, et al. Renal pathology patterns in type II diabetes mellitus: relationship with retinopathy. The Collaborative Study Group. *Nephrol Dial Transplant*, 1998; 13: 2547-52

12. Shimizu M, Furuichi K, Toyama T, et al. Long-term outcomes of Japanese type 2 diabetic patients with biopsy-proven diabetic nephropathy. *Diabetes Care*, 2013; 36: 3655-62