TRANSPLANT INTERNATIONAL

Transplant International ISSN 0934-0874

ORIGINAL ARTICLE

Pre-implantation analysis of kidney biopsies from expanded criteria donors: testing the accuracy of frozen section technique and the adequacy of their assessment by on-call pathologists

Amaia Sagasta,[1,*] Ana Sánchez-Escuredo,[2,*] Frederic Oppenheimer,[2] David Paredes,[3] Mireia Musquera,[4] Josep Maria Campistol[2] and Manel Solé[1]

1 Pathology Department, Hospital Clinic, Universitat de Barcelona, Barcelona, Spain
2 Nephrology and Renal Transplant Department, Hospital Clinic, Universitat de Barcelona, Barcelona, Spain
3 Transplant Service Fundation, Hospital Clinic, Universitat de Barcelona, Barcelona, Spain
4 Urology Department, Hospital Clinic, Universitat de Barcelona, Barcelona, Spain

Keywords
expanded criteria donor, frozen versus paraffin sections, interobserver agreement, kidney transplantation, pre-implantation biopsy.

Correspondence
Manel Solé, Pathology Department, Hospital Clinic, C/Villarroel 170, 08036-Barcelona, Spain.
Tel.: +34 932275400;
fax: +34 932275454;
e-mail: msole@clinic.ub.es

*Co-junior authors, both authors contributed equally.

Conflicts of interest
The authors have declared no conflicts of interest.

Received: 30 July 2015
Revision requested: 20 August 2015
Accepted: 23 October 2015
Published online: 20 November 2015

doi:10.1111/tri.12709

Summary

Pre-implantation renal biopsies of expanded criteria donors are one of the criteria used for allocation decisions, but there are concerns about the impact of the interobserver variability and the technique to be used. The aim was (i) to compare the original report performed by on-call pathologists using frozen sections (FS) to a retrospective analysis carried out by a trained pathologist using the same frozen section, and (ii) to compare the same FS to subsequently obtained paraffin sections (PS) by the same pathologist. A total of 92 biopsies, 78 from transplanted and 14 from nontransplanted cases, were analyzed. Agreement between observers using the same FS was weaker than the correlation between FS and PS in all the examined parameters (Kendall's Tau b for the Remuzzi score 0.104 vs. 0.306). According to the Remuzzi score, the revised FS analysis would have resulted in a higher rate of organ discard (n = 19) than PS (n = 14) and the original report (n = 6). However, kidneys that would have been discarded according to the retrospective analysis showed adequate outcomes in terms of graft survival and function. Accordingly, the impact of interobserver and technique-related variability can be minimized by the use of a relatively low threshold (RS $\leq$ 4) for organ acceptance.

Introduction

Controversy surrounds the utility of pre-implantation kidney biopsies as a tool for selecting grafts for expanded criteria donor (ECD) transplantation. Despite the widespread use of biopsies, there are no universally accepted practice guidelines on the best technique to be used or on the interpretation and reporting of the findings. Therefore, the value of pre-implantation kidney biopsy in predicting allograft survival is still debated in the literature [1–5].

In addition, interobserver variability is a well-known phenomenon that may influence the grading of histological lesions, thus compromising the final diagnosis and clinical decisions [6,7].

In daily practice, frozen section (FS) examination is used in many institutions for decision-making, but

reports of the reproducibility and prognostic value of donor biopsies are based on formalin-fixed and paraffin-embedded tissue. The superiority of paraffin over freezing has not been demonstrated for this particular purpose, and few studies have addressed this issue [8]. However, determination of the optimal technique is also important from the administrative and logistic points of view: the time required for diagnosis is at least 3 h even with the fastest paraffin section (PS) technique but can be less than 30 min with the FS method; moreover, the paraffin method requires resource use that in many centers would only be acceptable if this technique had proven superiority.

The aim of this study was to assess the reliability of FS evaluation by analyzing the correlation between: (i) distinct observers (on-call pathologists vs. a trained pathologist) when using FS (the technique currently used in our department for this purpose) and (ii) different techniques (PS vs. FS) when evaluated by the same observer. Both analyses included routinely used parameters and scores, as well as evaluation of the importance of the observed differences in (i) and (ii) in terms of organ acceptance and viability.

Methods

Donors and study design

This is an observational retrospective study assessing ECD pretransplantation biopsies. ECD are defined as donors over 60 years or between 50 and 59 years with at least two of the following conditions: hypertension history, serum creatinine >1.5 mg/dl or cause of death from cerebrovascular accident.

From the total pool of ECD pretransplantation wedge kidney biopsies at our institution between January 2000 and December 2008 (n = 424), grafts that did not include complete information about outcome were excluded (referred to other centers after transplantation and therefore lost for outcome analyses (n = 37)). Of the remaining 387 grafts, we randomly selected 100 grafts with available original report and frozen sections and paraffin block. Eight cases were subsequently discarded during the retrospective evaluation due to technical deficiencies in the original frozen sections (n = 5) or for exhausted tissue in the paraffin block (n = 3). Therefore, 92 specimens (for correlation between the two techniques) and 82 specimens (for correlation between observers) from wedge kidney biopsies were studied. These 92 specimens included 78 biopsies from transplanted specimens and 14 biopsies from nontransplanted specimens, which were blindly selected to reduce selection bias. The reasons for exclusion in this second group were a high (>4) Remuzzi score in the FS pre-implantation biopsy.

Extraction of clinical and pathological information from the patients' medical records and reporting of these data in this study were approved by the Ethics Review Board of Hospital Clinic in Barcelona.

Biopsy evaluation

Wedge biopsies obtained after removal of donor kidneys were sent to the pathology department for immediate pre-transplantation evaluation by FS and for processing for PS for the final evaluation.

All the biopsies included in this study were evaluated through the following three ways: (i) The original report of the FS at the time of transplantation, (ii) A second, retrospective evaluation of the FS by a pathologist trained specifically for this purpose, and (iii) A retrospective evaluation of the paraffin-embedded, PAS-stained permanent sections by the same trained pathologist.

The conditions and methods used in each evaluation are described below and summarized in Table 1:

(i) The evaluation at the time of transplantation was made by the on-call pathologist, who could be any of the 10 different pathologists in our department, 9 of whom were specialists in an area of pathology other than renal pathology and 1 of whom was a specialist in renal pathology. Clinical information available at the time of this scoring was donor age, gender, and risk factors. Three-micron-thick haematoxylin and eosin (H&E)-stained FS were evaluated. All sections were evaluated for the number of glomeruli and the percentage of global glomerulosclerosis, interstitial fibrosis, tubular atrophy, and fibrous intimal thickening.

Glomerulosclerosis was categorized into grades 0, 1, 2, and 3, corresponding to 0, 1–20%, 21–30%, and more than 30% global sclerosis, respectively. The degrees of interstitial fibrosis, tubular atrophy, and degree of vascular damage were graded from 0 to 3+ using Remuzzi score's definition (score of 0 if no changes were observed, score of up to 3 if marked changes were present -interstitial fibrosis score 3 when more than 50 percent of the renal parenchyma was replaced by connective tissue, tubular score 3 when more than 50 percent of tubules were atrophic and vascular score 3 when the vessel-wall thickness exceeded the luminal diameter or the lumen was occluded-) [9]. With the aim of improving reproducibility in the statistical analysis, a combined score of interstitial fibrosis and tubular atrophy was retrospectively obtained by selecting the highest score of the two items [10].

Thereafter, the biopsy specimens were fixed in formalin and embedded in paraffin blocks and included in our archives.

The scoring was recorded in the original report and later recovered for the study.

Table 1. Analyzed parameters, criteria, and scoring for the histologic parameters evaluated in the different conditions of the study.

Condition (i): Original report of the FS at the time of the transplantation		
Glomerulosclerosis	Grade 0	Absent
	Grade 1	1–20%
	Grade 2	21–30%
	Grade 3	>30%
Interstitial fibrosis	Grade 0	Absent
	Grade 1	<20% replacement by fibrous tissue
	Grade 2	20–50%
	Grade 3	>50%
Tubular atrophy	Grade 0	Absent
	Grade 1	<20% of tubuli affected
	Grade 2	20–50%
	Grade 3	>50%
Vascular damage	Grade 0	Absent
	Grade 1	Increased wall thickness, less than the luminal diameter
	Grade 2	Wall thickness equal or slightly greater then luminal diameter
	Grade 3	Wall thickness exceeds the luminal diameter
Total score [9]	Sum of previous four	0–12
Combined score of interstitial fibrosis and tubular atrophy [10]	Highest score of TA/IF	0–3
Condition (ii): retrospective evaluation of the FS by a pathologist trained specifically for this purpose		
(ii. 1) For interobserver agreement: same as in (i)		
(ii. 2) For agreement between techniques: same as in (i) plus arteriolar hyaline thickening [11]	Grade 0	Absent
	Grade 1	Mild to moderate in at least 1 arteriole
	Grade 2	Moderate to severe in >1 arteriole
	Grade 3	Severe in many arterioles
Condition (iii): retrospective evaluation of the paraffin-embedded, PAS-stained permanent sections by the same trained pathologist		
Same as in (ii), including arteriolar hyaline thickening		

FS, Frozen Sections; TA/IF, Tubular Atrophy/Interstitial Fibrosis.

(ii) The biopsies were retrospectively evaluated by the trained pathologist in charge of the study using the original FS slides. This evaluation was completely blind, as the pathologist was unaware of the original report, of the outcome of the evaluated kidney, and of any clinical information. The same items and scores were evaluated to create a new, comparable report. In addition, arteriolar hyaline thickening (defined as amorphous, homogeneous eosinophilic deposits in the arteriole wall) was graded from 0 to 3+ using definitions suggested by the Banff Schema of allograft pathology [11].

(iii) The biopsies were retrospectively evaluated once more by the trained pathologist, using paraffin-embedded, periodic acid-Schiff (PAS)-stained permanent sections. To do this, the original blocks were recovered and new three-micron-thick sections were made. The evaluation was performed blind, with the pathologist unaware of previous reports, and was made upon the same items and same scores as in (ii).

At the time of transplantation, the final score was calculated as the sum of the scores for the four compartments considered in the Remuzzi score (glomeruloesclerosis, tubular atrophy, interstitial fibrosis, and vascular damage) and therefore ranged from 0 (no lesions) to 12 points. In terms of histology results, the graft was accepted for implantation when the score was less than or equal to 4 points and excluded if more than 4 points.

Statistical analysis

Categorical variables are expressed as frequencies and percentages. Agreement between observers and techniques was evaluated using Kendall's Tau b correlation coefficient, a measure that is used to evaluate correlation when using ordinal variables and is corrected for agreement by chance [12]. The values range between -1 (perfect disagreement) and 1 (perfect agreement), a value of zero indicating the absence of association. The concordance between acceptance and discard for organ allocation was evaluated using the Kappa index, with values <0 indicating no agreement and 0–0.20 as slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1 as almost perfect agreement [13]. The statistical analysis was performed using SAS version 9.3 (SAS Institute Inc., SAS Campus Drive, Cary, NC, USA).

Results

Donor and recipient characteristics

We evaluated 92 kidneys from 50 donors (42 donors with left and right kidney and 8 donors with only one kidney).

The median age of donors was 65 $\pm$ 8 years, 41.6% were male, and the mean body mass index was 26.48 $\pm$ 3.14; 67.8% had high blood pressure, 24% were diabetics, and 15.7% showed acute renal failure at donation.

The median age of recipients was 60 $\pm$ 9 years. In all, 46.4% were male, the mean body mass index was 25.69 $\pm$ 3.72, cold ischemia time was 17.2 $\pm$ 5.2 h, 76.8% had high blood pressure, and 14.5% had diabetes mellitus. The mean length of stay was 16 $\pm$ 12 days, 30.4% had delayed graft function, and the mean creatinine level at discharge was 2.8 $\pm$ 1.7 mg/dl.

Sample characteristics: Histological lesions in frozen and paraffin sections

The median number of glomeruli in the evaluated specimens for FS and PS was 35 ± 18 and 46 ± 33, respectively. None of the samples included less than 10 glomeruli, and 89% of the evaluated samples included more than 20 glomeruli.

Table 2 shows the histological features of the samples: on retrospective review, interstitial fibrosis, tubular atrophy, and vascular lesions were scored as 1, followed by a score of 0 in most samples. For glomerulosclerosis, most of the samples were scored as 1 followed by scores 2 and 3, with few cases scored as 0. Arteriolar hyalinosis was the only parameter with a score 0 in most of the sections. The Remuzzi score was 4 points or less in 80 of the specimens.

Histological lesions in right and left kidneys from the same donor were of similar grade when evaluated in paraffin sections, the Remuzzi score being the same ± 1 point in 71.87% of the cases, although absolute concordance occurred only in 28.75% (Kendall's Tau b: 0.27). Concordance was lower in the retrospective review of frozen sections (Kendall Tau b for Remuzzi score: 0.03), and better in the original report (Kendall's Tau b for Remuzzi score: 0.67).

Interobserver correlation in frozen sections

The correlation between the on-call pathologists and the trained pathologist when evaluating the biopsies in FS was weak in all the parameters. Kendall's Tau b was 0.19 (95% CI: −0.06–0.45) for glomerulosclerosis, 0.10 (95% CI: −0.09–0.30) for tubular atrophy, 0.24 (95% CI: 0.03–0.44) for interstitial fibrosis, and 0.21 (95% CI: 0.005–0.41) for fibrous intimal thickening. Kendall's Tau b as an indicator of agreement in the Remuzzi score between the on-call pathologists and the trained pathologist was 0.10 (95% CI: −0.09–0.30). The trained pathologist assigned higher Remuzzi scores than the on-call pathologists (Table 3).

Similar results were found for organ acceptance: 14 and 6 grafts were excluded by the trained pathologist and on-call pathologists, respectively, indicating that the trained pathologist excluded 9.75% more organs than the on-call pathologists. Three cases were excluded by both the trained pathologist and the original report. Thus, the concordance in organ acceptance this group was fair (kappa index 0.33 (95% CI: 0.05–0.61)).

Correlation between techniques

The correlation between FS and PS showed Kendall's Tau b values of 0.36 (95% CI: 0.13–0.59) for glomerulosclerosis, which was the parameter with the best correlation. Values were similar for interstitial fibrosis (0.35 [95% CI: 0.15–0.55]), for intimal thickening (0.31 [95% CI: 0.11–0.51]), and for arteriolar hyalinosis (0.32 [95% CI: 0.18–0.46]). The parameter with the worst correlation was tubular atrophy, with a Kendall's Tau b of 0.16 (95% CI: −0.13–0.45). The agreement for Remuzzi score calculated by Kendall's Tau b was 0.31 (95% CI: 0.15–0.47). The trained pathologist assigned higher Remuzzi scores when using FS than when using PS (Table 4).

Similar results were found for organ acceptance, 19 and 12 grafts would hypothetically be excluded by FS and PS, respectively, indicating that the exclusion rate was 7.6% higher with frozen evaluation. Four cases were excluded by both FS and PS evaluation. Thus, the concordance in organ acceptance in this group was fair (kappa index 0.35 [95% CI: 0.11–0.59]).

Combined IFTA score

To improve interobserver agreement, we evaluated a combined score for interstitial fibrosis and tubular atrophy, obtained by selecting the highest scores of the two items.

However, the introduction of this combined tubulointerstitial scoring system did not improve interobserver or intra-observer agreement on analysis of these parameters (Tau b K for interstitial fibrosis 0.346; tubular atrophy 0.15; combined parameters (interstitial fibrosis/tubular atrophy) 0.157).

Table 2. Histological features of the samples when revised by the same observer (n = 92). The results are expressed by number of cases and percentage (N (%).

	Frozen section scores			Paraffin section scores		
Parameter	0	1	>1	0	1	>1
GS	6 (6.52)	73 (79.35)	13 (14.13)	8 (8.70)	71 (77.17)	13 (14.13)
IF	41 (44.57)	48 (52.17)	3 (3.26)	42 (45.65)	48 (52.17)	2 (2.17)
TA	16 (17.39)	73 (79.35)	3 (3.26)	8 (8.70)	82 (89.13)	2 (2.17)
CV	29 (34.12)	47 (55.29)	7 (8.24)	37 (43.53)	43 (50.59)	3 (3.53)
AH	87 (94.57)	5 (5.43)	0 (0)	60 (65.22)	31 (33.70)	1 (1.09)

GS, glomerulosclerosis; IF, interstitial fibrosis; TA, tubular atrophy; CV, fibrous intimal thickening; AH, arteriolar hyalinosis.

Table 3. Distribution of the Remuzzi score for frozen sections in the original report and frozen sections evaluated by the trained pathologist (n = 82). The results are expressed as the number of cases and percentage (N [%]).

	Remuzzi score								
	0	1	2	3	4	5	6	7	8
OR	3 (3.66)	20 (24.39)	15 (18.29)	21 (25.61)	17 (20.73)	3 (3.66)	2 (2.44)	0 (0)	1 (1.22)
TP FS	0 (0)	9 (10.98)	16 (19.51)	23 (28.05)	20 (24.39)	9 (10.98)	3 (3.66)	2 (2.44)	0 (0)

OR: original report, frozen section; TP FS: trained pathologist, frozen section.

Table 4. Distribution of Remuzzi score for FS and PS evaluation revised by the same observer (n = 92). The results are expressed as the number of cases and percentage (N [%]).

	Remuzzi score								
	0	1	2	3	4	5	6	7	8
FS	0 (0)	10 (10.87)	18 (19.57)	23 (25)	22 (23.91)	12 (13.04)	4 (4.35)	2 (2.17)	1 (1.09)
PS	2 (2.17)	8 (8.7)	17 (18.48)	27 (29.35)	26 (28.26)	8 (8.7)	3 (3.26)	1 (1.09)	0 (0)

FS, frozen section; PS, paraffin section.

Outcomes

Patient survival and death-censored kidney survival in the whole series was 88.1% (n = 70) and 91.2% (n = 73), respectively, at 1 year and 68.3% (n = 54) and 73.1% (n = 58) at 5 years. The mean serum creatinine level was 1.8 ± 0.6 mg/dl at 1 year and was 2.2 ± 1.2 mg/dl at 5 years.

Table 5 shows the outcomes of transplanted patients with discordant results between the original report and the evaluation by the trained pathologist. Fourteen of the grafts were excluded according to the retrospective evaluation by the trained pathologist but were eligible according to the original report, and thus, they were transplanted; of them, only one was lost due to graft thrombosis. The rest achieved adequate graft survival and function, with mean serum creatinine level of 1.9 ± 0.6 mg/dl at 1 year and 2.5 ± 1 mg/dl at 5 years. Renal function at 1 year in this group was not significantly worse than for the whole transplanted group (Cr 1.93 ± 0.7 mg/dl) and for the group with revised S ≤ 4 Cr 1.8 ± 0.62 (P 0.127).

Table 5. Outcome of transplanted organs in discordant cases between the original report and the trained pathologist evaluation using frozen and paraffin sections.

	Cr at 1 year	Cr at 5 years	Graft survival censored by death	Patient survival
1	2, 7	3, 5	Yes	Yes
2	1, 4	1, 7	Yes	Yes
3	1, 9	3, 3	Yes	Yes
4	2	2, 2	Yes	Yes
5	0, 6	NA	Yes	Exitus (Tumor at 6 m)
6	1, 9	1, 8	Yes	Yes
7	PNF	PNF	Thrombosis	Yes
8	2	NA	Yes	Yes
9	1, 9	1, 7	Yes	Yes
10	2, 5	NA	Yes	Exitus (Infection at 18 m)
11	1, 5	NA	Yes	Exitus (cardiovascular 16 m)
12	1, 5	1, 4	Yes	Yes
13	1, 7	2, 4	Yes	Yes
14	2, 9	4, 5	Yes	Yes

Cr, creatinine in mg/dl; m, months; PNF, primary nonfunction; NA, not available.

Discussion

Our study provides new data on the histological evaluation of pre-implantation kidney transplant biopsies that were used for the allocation of ECD kidneys. The original evaluations carried out by on-call pathologists using FS were compared to a retrospective analysis carried out by a trained pathologist. This comparison revealed that (i) the trained pathologist assigned higher RSs than the on-call pathologists and, accordingly, (ii) renal transplant acceptance was higher by the on-call pathologists than by the trained pathologist. We also compared the concordance between FS and PS evaluation, revealing that (i) the trained pathologist assigned higher RSs when using FS than with PS and, accordingly (ii) hypothetical organ acceptance comparing both techniques excluded more organs from transplantation when using FS than using PS. Additionally, these results allow us compare concordance between observers and concordance between techniques. So, the overall results indicate that regarding concordance in the evaluation of pre-implantational biopsies for selection purposes,

interobserver variability produces more discordances than the technique used for sample processing, but in any case these discordances had no significant impact on outcomes for the transplanted organs. In these terms, the Remuzzi score, which is used to assess graft eligibility, was the parameter with better improvement in correlation (Remuzzi score for interobserver agreement 0.104; Remuzzi score for agreement between techniques 0.306). This result would suggest that specific training is irrespective of the technique used and that to improve concordance, staff training may be more useful than replacing the frozen technique by paraffin processing of these samples. However, the need for specific training or expertise for this kind of evaluation is controversial. Whereas in the study by Furness *et al.* [6] and in spite of the complexity of the evaluated parameters, attempts to improve reproducibility failed to produce any detectable improvement, in the report by Azancot *et al.* [7] the retrospective review by a renal pathologist detected grafts with poor function better than the scores provided by the on-call nonspecialized pathologists. By contrast, our study indicates that minor changes detected by an intentional retrospective review may increment the total score in several points, but with no significant impact on outcomes, favoring that the original selection by the on-call pathologist is efficient enough.

Crude comparisons between the original report and revised scores are biased by the perceived relevance of decisions taken by on-call pathologists. The analysis of concordance between right and left kidney evaluations in our study highlights this potential bias: concordance figures were much better in the original report (Kendall's Tau b for RS: 0.67) than in any of the retrospective reviews (Kendall's Tau b for RS in FS: 0.03 and in PS: 0.27). In a similar way, on-call pathologists would overlook minor changes in a particular item when total scores are clearly above or below the cut-offs for organ acceptance. Thus, a blind evaluation in a retrospective analyses would result in more pronounced discordances than expected. This approach would also compromise the validity of comparing the predictive value of these scores with outcomes [7].

In our study, variability detected in the individual parameters are mostly in the 0 to 1 range (Table 2), indicating that the original evaluation by the on-call pathologists effectively detected organs with severe lesions. Therefore, although correlation coefficients seem to be poor for most parameters as well as for the total scores, this, again, has no impact in clinical terms. Nevertheless, an adequate reproducibility of the evaluation of the individual parameters is necessary for the success of any scoring system. It has been suggested that evaluation of the parameters analyzed in FS is more difficult and would probably result in greater interobserver variation than the use of formalin-fixed sections [6] but there are not previously published analysis based on FS. The results of our study showed that concordance between individual parameters when comparing FS and PS was best for glomerulosclerosis. This finding is in agreement with previous reports on PS, as many groups consider that glomerulosclerosis is the easiest and most helpful item in scoring [2,14,15]. However, interobserver concordance for glomerulosclerosis was lower than expected and below values for interstitial fibrosis and fibrous intimal thickening. The parameter with the worst correlation in both study groups was tubular atrophy. In relation to the differences between techniques, we consider tubular atrophy to be a particularly difficult parameter to assess in FS due to tissue retraction after processing, so that normal tubules may be more easily misinterpreted as atrophic. For interobserver differences, the criterion used to assess this particular item was grade 0 for "no tubular atrophy", as suggested by the Banff Schema of allograft pathology [11]. However, this statement is difficult to assess, as individual atrophic tubules can be easily missed and/or mistaken, especially in FS, which would change the score from 0 to 1. Trying to improve the reproducibility for FS by combining interstitial fibrosis with tubular atrophy, as is used in the IgA nephropathy scoring system [16] and suggested by Snoeijs *et al.* [10], did not improve the correlation in this case. The evaluation of arteriolar hyaline thickening, a feature considered relevant for outcome [1,15], seems inadequate for FS, although the incidence of this feature in our series was relatively low.

Additionally, and given the good outcomes of organs with revised score higher than 4, it would be interesting to determine if the threshold should be modified in order toincrease the rate of organ acceptance. This is difficult to address from our results as if we take the PS score as the gold standard, the original report would be underscoring, whereas the trained pathologist on FS would be overscoring. In the Remuzzi score, as in the Banff based scores, thresholds for individual parameters and for organ acceptance are arbitrary. In spite of that, RS has demonstrated to be adequate for organ selection. In a recent analysis on basal biopsies, the Remuzzi score was the only independent predictor for kidney graft survival [5]. Therefore, and at least for frozen section analysis, it seems wise to maintain a low threshold, while basing the eventual acceptance of organs with borderline RS scores (i.e. RS 5 or 6) on additional criteria, such as perfusion machine results, when available, or clinical scores such as the KDPI index [17].

A limitation to this type of study, rather than to our study itself, is that the examined biopsies have been previously selected by pathological analysis on FS or perfusion data, with kidneys with greater degrees of injury being excluded. In these samples, the disagreements may be more difficult to identify and may not be clinically relevant. To minimize this bias, we blindly included biopsies from

kidneys excluded at the time of the original report in the study group.

Equally, a major drawback when analyzing differences between techniques is the impossibility of using the same tissue section. To overcome this limitation, comparable representativity of the tissue samples must be assumed. However, interobserver variability was evaluated using the same tissue sections and still showed the worst values. This could indicate that, despite being an undeniable limitation, the use of different tissue sections seems to have little influence on the results.

The purpose of biopsies on ECD is to identify good organs within a pool of individuals with risk factors that, otherwise, would exclude them as donors. The main conclusion of our study is that frozen section analysis by non-specialized pathologists using the RS is safe, as even those cases that were underscored according to the retrospective review by a specifically trained pathologist achieved adequate function in the follow-up. The impact of interobserver variability is minimized by the use of a relatively low threshold ($RS \leq 4$) that selects organs with only mild lesions. Decisions about organs with higher scores (RS 5 or 6) should be based on other criteria regarding the organ (i.e. resistance and flux indices), clinical data of the donor (i.e. KDPI index) or the recipient's condition.

Authorship

AS: collected, analyzed histological data and wrote the manuscript. AS-E: collected, analyzed clinical data and wrote the manuscript. FO and MS: design of the study and manuscript review. DP and MM: data collection. JMC: design of the study.

Funding

The authors have declared no funding.

Acknowledgements

The authors would like to thank Abiguei Torres for statistic analyses.

References

1. El-Husseini A, Sabry A, Zahran A, *et al.* Can donor implantation renal biopsy predict long-term renal allograft outcome? *Am J Nephrol* 2007; **27**: 144.
2. Serón D, Anaya F, Marcén R, *et al.* Guidelines for indicating, obtaining, processing and evaluating kidney biopsies. *Nefrologia* 2008; **28**: 385.
3. Navarro MD, López-Andréu M, Rodríguez-Benot A, *et al.* Significance of preimplantation analysis of kidney biopsies from expanded criteria donors in long-term outcome. *Transplantation* 2011; **91**: 432.
4. De Vusser K, Lerut E, Kuypers D, *et al.* The predictive value of kidney allograft baseline biopsies for long-term graft survival. *J Am Soc Nephrol* 2013; **24**: 1913.
5. Kosmoliaptsis V, Salji M, Bardsley V, *et al.* Baseline donor chronic renal injury confers the same transplant survival disadvantage for DCD and DBD kidneys. *Am J Transplant* 2015; **15**: 754.
6. Furness PN, Taub N, Assmann KJ, *et al.* International variation in histologic grading is large, and persistent feedback does not improve reproducibility. *Am J Surg Pathol* 2003; **27**: 805.
7. Azancot MA, Moreso F, Salcedo M, *et al.* The reproducibility and predictive value on outcome of renal biopsies from expanded criteria donors. *Kidney Int* 2014; **85**: 1161.
8. Goumenos DS, Kalliakmani P, Tsamandas AC, *et al.* The prognostic value of frozen section preimplantation graft biopsy in the outcome of renal transplantation. *Ren Fail* 2010; **32**: 434.
9. Remuzzi G, Cravedi P, Perna A, *et al.* Dual Kidney Transplant Group. Long-term outcome of renal transplantation from older donors. *N Engl J Med* 2006; **354**: 343.
10. Snoeijs MG, Boonstra LA, Buurman WA, *et al.* Histological assessment of pre-transplant kidney biopsies is reproducible and representative. *Histopathology* 2010; **56**: 198.
11. Solez K, Colvin RB, Racusen LC, *et al.* Banff 07 classification of renal allograft pathology: updates and future directions. *Am J Transplant* 2008; **8**: 753.
12. Kendall MG. *Rank Correlation Methods*, 2nd ed. London: Charles Griffin & Co., Ltd, 1955.
13. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* 1968; **70**: 213.
14. Anglicheau D, Loupy A, Lefaucheur C, *et al.* A simple clinico-histopathological composite scoring system is highly predictive of graft outcomes in marginal donors. *Am J Transplant* 2008; **8**: 2325.
15. Munivenkatappa RB, Schweitzer EJ, Papadimitriou JC, *et al.* The Maryland aggregate pathology index: a deceased donor kidney biopsy scoring system for predicting graft failure. *Am J Transplant* 2008; **8**: 2316.
16. Roberts IS, Cook HT, Troyanov S, *et al.* The Oxford classification of IgA nephropathy: pathology definitions, correlations, and reproducibility. *Kidney Int* 2009; **76**: 546.
17. Gandolfini I, Buzio C, Zanelli P, *et al.* The Kidney Donor Profile Index (KDPI) of marginal donors allocated by standardized pretransplant donor biopsy assessment: distribution and association with graft outcomes. *Am J Transplant* 2014; **14**: 2515.