see commentary on page 1016

# The reproducibility and predictive value on outcome of renal biopsies from expanded criteria donors

M. Antonieta Azancot[1], Francesc Moreso[1], Maite Salcedo[2], Carme Cantarell[1], Manel Perello[1], Irina B. Torres[1], Angeles Montero[2], Enric Trilla[3], Joana Sellarés[1], Joan Morote[3] and Daniel Seron[1]

*[1]Department of Nephrology, Hospital Universitari Vall d'Hebron, Universitat Autònoma Barcelona, Barcelona, Spain; [2]Department of Pathology, Hospital Universitari Vall d'Hebron, Universitat Autònoma Barcelona, Barcelona, Spain and [3]Department of Urology, Hospital Universitari Vall d'Hebron, Universitat Autònoma Barcelona, Barcelona, Spain*

**Reproducibility and predictive value on outcome are the main criteria to evaluate the utility of histological scores. Here we analyze the reproducibility of donor biopsy assessment by different on-call pathologists and the retrospective evaluation by a single renal pathologist blinded to clinical outcomes. We also evaluate the predictive value on graft outcome of both evaluations. A biopsy was performed in donors with any of the following: age $\geqslant$55 years, hypertension, diabetes, creatinine $>$1.5 mg/dl, or stroke. Glomerulosclerosis, interstitial fibrosis, tubular atrophy, intimal thickening, and arteriolar hyalinosis evaluated according to the Banff criteria were added to obtain a chronic score. Biopsies were classified as mild ($\geqslant$3), intermediate (4–5), or advanced (6–7) damage, and unacceptable ($\geqslant$8) for transplantation of 127 kidneys biopsied. Weighted κ value between both readings was 0.41 (95% CI: 0.28–0.54). Evaluation of biopsies by the renal pathologist was significantly and independently associated with estimated 12-month glomerular filtration rate and a significant composite outcome variable, including death-censored graft survival and time to reach an estimated glomerular filtration rate $<$30 ml/min per 1.73 m$^2$. Thus, there was no association between readings of on-call pathologists and outcome. The lack of association between histological scores obtained by the on-call pathologists and graft outcome suggests that a specific training on renal pathology is recommended to optimize the use of kidneys retrieved from expanded criteria donors.**

*Kidney International* (2014) **85,** 1161–1168; doi:10.1038/ki.2013.461; published online 27 November 2013

KEYWORDS: donor pre-implantation biopsies; expanded criteria donors; kidney transplantation

**Correspondence:** *Francesc Moreso, Department of Nephrology, Hospital Universitari Vall d'Hebron, Passeig Vall d'Hebron 119-129, Barcelona 08035, Spain. E-mail: fjmoreso@vhebron.net*

**Received 9 November 2012; revised 29 August 2013; accepted 5 September 2013; published online 27 November 2013**

Renal transplantation is the best available treatment option for patients with end-stage renal disease as it offers a longer survival and a better quality of life at a lower cost than dialysis.[1,2] Despite the sustained effort to increase the number of kidneys retrieved from deceased and living donors, the number of patients in the waiting list is still growing.[3,4]

A steady increase in donor age has been paralleled by a rising number of discarded kidneys for transplantation.[3,5] Different policies have been proposed to safely transplant kidney from aged donors or donors with comorbidities, the so called expanded criteria donors (ECDs),[6] as they offer a survival benefit in comparison to patients who remain on the waiting list.[7,8] These kidneys are usually offered to aged recipients in order to match kidney and patient survival expectancies according to an old-for-old policy.[9,10] Sophisticated statistical models evaluating donor and recipient characteristics have been proposed to better estimate risk for graft failure based on clinical data, but these proposals have not been widely used.[11–13] Instead, many centers evaluate donor biopsies to assess the risk of graft failure.[14] The consideration of histological information to decide the acceptance of a kidney for transplantation is based on different reports showing that donor histological damage is associated with renal function and graft survival.[15–19] In many centers, the decision to accept or discard a kidney on a daily basis practice relies on the histological evaluation by different on-call pathologists. Interobserver variability in the evaluation of biopsies is an important issue as it may influence grading of histological lesions and consequently affect the final diagnosis and clinical decisions.[20]

At our center, we perform donor biopsies as a standard of care since January 2009 in donors at risk for graft failure. Biopsies are paraffin embedded and graded by the pathologist on-call to decide whether the kidney is suitable for implantation. In the present study, we evaluate the agreement on the scoring of donor biopsies between the on-call pathologists and the retrospective evaluation by the renal pathologist on the same sections. Additionally, we study the predictive value on outcome of both biopsy scores.

## RESULTS

### Donor and recipient characteristics

Between January 2009 and May 2011, 182 kidneys from deceased donors were evaluated and recipients were followed up until April 2013. Twenty-one kidneys (11.5%) were discarded owing to the following reasons: donor neoplasia ($n=3$), multiple pyelonephritic scars ($n=2$), severe atherosclerosis of the renal arteries ($n=3$), giant cortical cyst ($n=1$), histology with disseminated intravascular coagulation ($n=2$), and a donor score $\geqslant 8$ in both kidneys ($n=10$). Thus, we evaluated 161 transplants, 44 (27%) harvested from standard criteria donors and 117 (73%) from ECDs. Among ECDs, there were 91 (77.8%) aged $\geqslant 55$ years, 79 (67.5%) with history of arterial hypertension, 74 (63.2%) deceased due to stroke, 26 (22.3%) with history of diabetes mellitus, and 5 (3.7%) with final pre-procurement serum creatinine $>1.5$ mg/dl. In 82 cases (70.0%), at least two of the above-mentioned risks factors were present. Six out of the 8 transplants performed with kidneys obtained from donors after cardiac death accomplished expanded criteria and were biopsied before implantation. There were two kidneys from two different donors that were re-biopsied owing to insufficient histological sample size and a discrepancy score $>3$ between both kidneys.

There was no difference between standard and ECD in the incidence of delayed graft function (29.5 vs. 24.8%) and acute rejection (18.2 vs. 17.1%). At 12 months, estimated glomerular filtration rate (eGFR) was higher in patients receiving a graft from a standard donor ($58\pm18$ vs. $46\pm16$ ml/min; $P=0.001$). During a mean follow-up of $35\pm16$ months, there were no significant differences between standard criteria and ECD in patient survival (93.5 vs. 91.8%), death-censored graft survival (92.4 vs. 88.6%), and graft survival (86.4 vs. 81.3%).

Eight out of the 117 ECD transplant patients died due to urinary sepsis ($n=2$), lung neoplasia ($n=2$), *Pseudomona aeruginosa* pneumonia ($n=1$), miliary tuberculosis ($n=1$), colitis due to cytomegalovirus and *Clostridium difficile* infection ($n=1$), and sudden death ($n=1$). Graft failure occurred in 10 patients owing to arterial thrombosis ($n=5$), urinary fistula ($n=1$), BK virus nephropathy ($n=1$), unrecovered acute renal failure associated with biliary sepsis ($n=1$), acute vascular rejection ($n=1$), and chronic humoral rejection ($n=1$).

### Reproducibility of histological scores

There were 127 donor biopsies for evaluation, 117 from transplanted and 10 from discarded kidneys. There were 122 out of 127 donor biopsies available for retrospective evaluation by the renal pathologist as five of them were processed and evaluated at another center. Biopsies were evaluated by 12 different on-call pathologists, and the median number of biopsies evaluated by each one was 12 (interquartile range 8–14). Mean number of glomerular sections per biopsy was $66\pm36$. There was a high correlation between percentage of glomeruloesclerosis evaluated by the on-call pathologists and the renal pathologist ($R=0.96$; $P<0.001$). Weighted κ, a statistical measure of agreement between the on-call and renal pathologist, for categorized glomerulosclerosis was 0.86 (95% confidence interval (CI): 0.77–0.95), and the Bland–Altman plot showed a bias of 0.4% and a coefficient of repeatability of $\pm 4\%$.

Weighted κ between on-call pathologists and the renal pathologist was 0.37 (95% CI: 0.22–0.51) for intimal thickening and 0.31 (95% CI: 0.15–0.46) for interstitial fibrosis. Lower κ values were obtained for arteriolar hyaline changes (0.25; 95% CI: 0.10–0.39) and for tubular atrophy (0.14; 95% CI: $-0.06$–0.34). Classification of donor damage as mild, intermediate, advanced, and unacceptable by the on-call pathologists and the renal pathologist is shown in Table 1. There were 39 biopsies (32.0%) classified as advanced damage (donor score 6-7) by the on-call pathologists while this figure was reduced to 23 (18.8%) when the renal pathologist read the biopsies ($P<0.001$). Weighted κ between both observations was 0.41 (95% CI: 0.28–0.54). Notably, according to the retrospective evaluation of the renal pathologist, 2 out of 10 discarded kidneys for donor score $\geqslant 8$ were considered adequate for transplantation. This represents that 10 out of 127 (7.8%) kidneys were discarded according to on-call pathologists and 8 out of 122 (6.5%) according to the renal pathologist. Donor damage score evaluated by on-call pathologists and renal pathologist showed a bias of 0.2 and a coefficient of repeatability of $\pm 2.89$ by Bland–Altman analysis. Lin's correlation coefficient of the donor score between on-call pathologists and the renal pathologist was 0.61 (95% CI = 0.48–0.71). The total deviation index (TDI) was 2.45 and the coverage probability for 1 and 2 points difference between observers were 0.49 and 0.82, respectively.

### Sources of disagreement in donor biopsy evaluation

There were 3 on-call pathologists with special training in renal transplant pathology who evaluated 20 biopsies and 9 without special training who evaluated 102 biopsies. Classification of donor damage as mild, intermediate, advanced, and

**Table 1 | Relationship between donor biopsy scores obtained by on-call pathologists and the renal pathologist (MS)**

| | | On-call pathologist | | | |
|---|---|---|---|---|---|
| Renal pathologist | Total | Mild damage ⩽3 | Moderate damage 4–5 | Advanced damage 6–7 | Unacceptable ⩾8 |
| Mild damage ⩽3 | 36 | 17 | 15 | 4 | 0 |
| Moderate damage 4–5 | 53 | 3 | 31 | 19 | 0 |
| Advanced damage 6–7 | 25 | 3 | 4 | 16 | 2 |
| Unacceptable ⩾8 | 8 | 0 | 0 | 0 | 8 |
| Total | 122 | 23 | 50 | 39 | 10 |

*P*-value $<0.001$ ($\chi^2$).

unacceptable for transplantation yielded a weighted κ for trained on-call pathologists of 0.50 (95% CI: 0.18–0.82) and 0.36 (95% CI: 0.20–0.53) for untrained. Intra-observer reproducibility of the renal pathologist was evaluated in a subset of 30 biopsies and weighted κ was 0.50 (0.27–0.74).

### Predictive value of donor score prospectively evaluated by the on-call pathologists and graft outcome

The following variables were different between mild, intermediate, and advanced scores: donor age ($P=0.001$), recipient age ($P=0.0024$), and donor hypertension ($P=0.001$), whereas the incidence of delayed graft function, the incidence of acute rejection, graft survival, death-censored graft survival, and patient survival were not different among groups (Table 2).

eGFR at 12 months (Figure 1a) and the temporal evolution of the composite outcome variable consisting on death-censored graft failure or time to reach an eGFR < 30 ml/min per 1.73 $m^2$ (Figure 2a) were no different among groups.

### Predictive value of donor biopsy score retrospectively evaluated by the renal pathologist and graft outcome

The following variables were different between mild, intermediate, and advanced scores: donor age ($P=0.0002$), recipient age ($P=0.0001$), and prevalence of donor hypertension ($P=0.0017$), whereas the incidence of delayed graft function, the incidence of acute rejection, graft survival, death-censored graft survival, and patient survival were not different among groups (Table 3).

Recipient 12-month eGFR was significantly decreased in patients with advanced histological scores (Figure 1b). eGFR at 12 months was also associated with donor age (R = -0.27, $P=0.0055$), stroke as the cause of donor's death (51 ± 14 vs. 43 ± 16 ml/min, $P=0.0099$), delayed graft function (48 ± 15 vs. 36 ± 14, $P=0.0027$) and acute rejection (47 ± 16 vs. 39 ± 12 ml/min, $P=0.0573$). Stepwise regression analysis showed that donor damage score obtained by the renal pathologist, delayed graft function and acute rejection were independent predictors of eGFR at 12 months ($R=0.49$, $P<0.0001$).

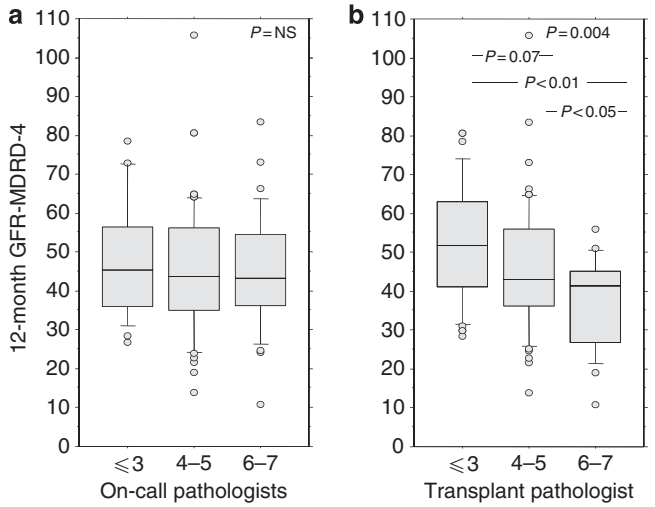


**Figure 1 | Box plot of 12 month estimated glomerular filtration rate (eGFR).** Box plot of 12 month eGFR according to donor damage score evaluated by (**a**) on-call pathologists and (**b**) renal pathologist. The line represents the median value of eGFR, the top and the bottom of the box are the 75th and 25th percentile and the ends of the whiskers represent the 90th and 10th percentile. Any data not included between the whiskers are plotted as an outlier represented by a circle. MDRD-4, Modification of Diet in Renal Disease 4. NS, not significant.

**Table 2 | Donor and recipient characteristics, 12-month eGFR, and graft survival in patients receiving an expanded criteria donor and classified according to the on-call pathologist donor score**

| Variable | Donor score ≤3 ($n=25$) | Donor score 4–5 ($n=52$) | Donor score 6–7 ($n=40$) | *P*-value |
|---|---|---|---|---|
| Donor age (years) | 53 ± 13 | 62 ± 9[a] | 62 ± 9[a] | <0.001 |
| Donor gender (male/female) | 14/11 | 28/24 | 29/11 | NS |
| Donor serum creatinine (mg/dl) | 1.1 ± 0.7 | 0.9 ± 0.3 | 0.9 ± 0.5 | NS |
| Donor hypertension (no/yes) | 15/10 | 18/34 | 4/36 | <0.001 |
| Donor diabetes (no/yes) | 23/2 | 38/14 | 30/10 | NS |
| Cause of donor death (trauma/stroke/DACD/other) | 7/14/2/2 | 12/32/2/4 | 7/28/2/3 | NS |
| Patient age (years) | 53 ± 16 | 61 ± 10[a] | 62 ± 9[a] | 0.0024 |
| Patient gender (male/female) | 18/7 | 30/22 | 28/12 | NS |
| Primary renal disease (GN/CTIN/ADPKD/DN/vascular/unknown/other) | 5/4/5/1/3/6/1 | 14/6/6/5/7/12/2 | 3/3/9/5/4/13/3 | NS |
| HLA mismatches | 4.2 ± 0.8 | 4.1 ± 1.0 | 4.0 ± 0.9 | NS |
| Transplant (primary/re-transplant) | 22/3 | 44/8 | 37/3 | NS |
| Cold ischemia time (hours) | 16 ± 4 | 16 ± 4 | 16 ± 4 | NS |
| Delayed graft function (no/yes) (%) | 20/5 (20%) | 41/11 (21.1%) | 28/12 (30%) | NS |
| Acute rejection (no/yes) (%) | 19/6 (24%) | 44/8 (15.4%) | 34/6 (15.0%) | NS |
| Patient survival | 100 | 92.7 | 86.8 | NS |
| Death-censored graft survival (%) | 83.6 | 90.1 | 92.2 | NS |
| Graft survival (%) | 83.6 | 83.4 | 80.0 | NS |

Abbreviations: ADPKD, autosomal dominant polycystic kidney disease; CTIN, chronic tubule-interstitial nephritis; DACD, donor after cardiac death; DN, diabetic nephropathy; eGFR, estimated glomerular filtration rate; GN, glomerulonephritis.
[a]*P*-value <0.05 versus donor score ≤3 by Scheffé's test.

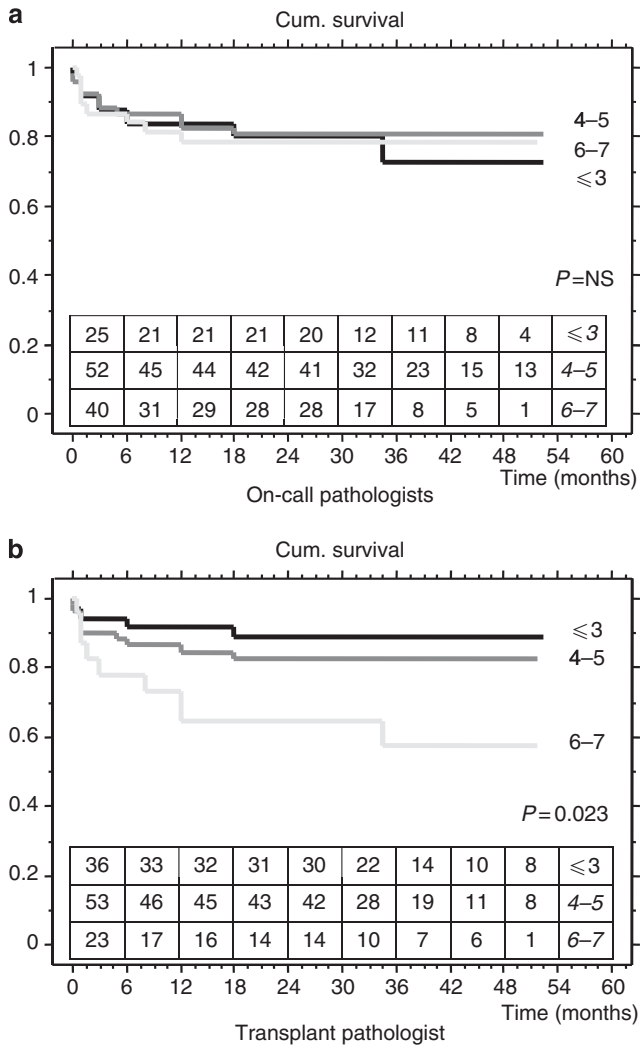


**Figure 2 | Kaplan–Meier survival curves for the composite outcome variable.** Kaplan–Meier survival curves for the composite outcome variable according to donor damage score evaluated by (**a**) on-call pathologists and (**b**) renal pathologist. The number of patients at risk at each time period is shown in the table.

Time to reach the composite outcome variable according to biopsy score was also significantly different among groups ($P = 0.023$) as shown in Figure 2b. Univariate Cox regression analysis showed that patients receiving a kidney with advanced damage had a higher risk to reach the composite variable than patients with mild damage (hazard rate (HR) 4.0, 95% CI: 1.2–13.0; $P = 0.021$) and patients with intermediate damage (HR: 2.4, 95% CI:0.97–6.35; $P = 0.059$). Stroke as the cause of donor death (HR: 2.8 and 95% CI: 1.2–6.9; $P = 0.022$) and acute rejection (HR: 2.9 and 95% CI: 1.4–6.1; $P = 0.004$) were also associated with the composite outcome variable. Multivariate Cox regression analysis showed that donor score evaluated by the renal pathologist was the only independent predictor of the composite outcome variable. This analysis showed that patients receiving a kidney with advanced damage had a higher risk to reach the composite variable than patients with mild damage (HR 3.6, 95% CI: 1.1–12.0; $P = 0.039$) and patients with intermediate damage (HR: 2.4, 95% CI: 0.9–6.1; $P = 0.064$).

### Reanalysis of data considering ECD according to the Organ Procurement and Transplantation Network/United Network for Organ Sharing (OPTN/UNOS) definition

We reanalyzed our data only considering the 90 kidneys accomplishing expanded criteria according to the OPTN/UNOS (donor age $\geqslant 60$ years or donor age $\geqslant 50$ years with at least two of the following: serum creatinine $> 1.5$ mg/dl, hypertension, or stroke as the cause of death).[9] This analysis confirmed an association between the donor damage score evaluated by the renal pathologist and 12-month eGFR ($49 \pm 15$, $46 \pm 18$, and $37 \pm 12$ ml/min per 1.73 $m^2$, for mild, moderate, and advanced damage score, respectively, $P = 0.047$). On the contrary, when this analysis was performed considering the donor damage score evaluated by on-call pathologists no associations were found ($40 \pm 10$, $45 \pm 18$, and $45 \pm 15$ ml/min per 1.73 $m^2$ for mild, moderate, and advanced damage score, respectively, $P =$ NS).

## DISCUSSION

The main finding of this study is the low reproducibility of the evaluation of donor renal biopsies between on-call pathologists and the renal pathologist. Furthermore, the scoring of donor biopsies by different on-call pathologists was not associated with graft outcome. On the contrary, the retrospective scoring of the same slides by the renal pathologist showed that donor pre-existing damage was an independent predictor of 12-month renal function as well as the outcome evaluated by means of a composite variable, including death-censored graft survival and time to reach an eGFR $< 30$ ml/min per 1.73 $m^2$.

Criteria to evaluate the utility of any histological score are predictive value on outcome and reproducibility. In fact, these two criteria were used to modify the Banff classification for acute rejection.[21,22] Our study shows that donor biopsies are not associated with graft outcome when evaluated by different on-call, non-renal trained pathologists. For this study, different statistical methods to assess agreement between on-call pathologists and the renal pathologist showed a low reproducibility between both observations.

Low reproducibility in renal allograft scoring between different observers has been previously described. In the CERTPAP study[20] evaluating 55 renal allograft biopsies that were circulated around 22 European centers, κ values for the evaluation of different lesions according to the Banff criteria were low. The main conclusion of this study was that in multicenter trials, including assessment of biopsies, histological evaluation should be carried out in a single center by a single pathologist. Although this routine has been systematically applied in clinical trials, it is difficult to apply it in the evaluation of donor biopsies, as different on-call pathologists are necessary to evaluate biopsies during a 24-h

**Table 3 | Donor and recipient characteristics, 12-month eGFR, and graft survival in patients receiving an expanded criteria donor and classified according to the transplant pathologist donor score**

| Variable | Donor score ⩽3 ($n=36$) | Donor score 4–5 ($n=53$) | Donor score 6–7 ($n=23$) | *P*-value |
|---|---|---|---|---|
| Donor age (years) | 54 ± 11 | 63 ± 9[a] | 63 ± 10[a] | <0.001 |
| Donor gender (male/female) | 20/16 | 35/18 | 12/11 | NS |
| Donor serum creatinine (mg/dl) | 1.1 ± 0.7 | 0.9 ± 0.3 | 0.9 ± 0.4 | NS |
| Donor hypertension (no/yes) | 19/17 | 11/42 | 4/19 | <0.001 |
| Donor diabetes (no/yes) | 32/4 | 37/16 | 17/6 | NS |
| Cause of donor death (trauma/stroke/DACD/other) | 11/18/3/4 | 10/37/1/5 | 5/17/0/1 | NS |
| Patient age (years) | 53 ± 16 | 61 ± 10[a] | 66 ± 8[a,b] | <0.001 |
| Patient gender (male/female) | 21/15 | 36/17 | 15/8 | NS |
| Primary renal disease (GN/CTIN/ADPKD/DN/vascular/unknown/other) | 5/5/7/4/4/10/1 | 11/6/7/6/7/12/4 | 5/2/5/0/2/8/1 | NS |
| HLA mismatches | 3.9 ± 1.1 | 4.2 ± 0.9 | 4.2 ± 0.8 | NS |
| Transplant (primary/re-transplant) | 31/5 | 47/6 | 20/3 | NS |
| Cold ischemia time (hours) | 15 ± 4 | 17 ± 4 | 16 ± 3 | NS |
| Delayed graft function (no/yes) (%) | 30/6 (16.7%) | 37/16 (30.2%) | 17/6 (26.1%) | NS |
| Acute rejection (no/yes) (%) | 28/8 (22.2%) | 46/7 (13.2%) | 20/3 (13.0%) | NS |
| Patient survival | 97.1 | 91.9 | 90.5 | NS |
| Death censored graft survival (%) | 81.3 | 94.5 | 84.5 | NS |
| Graft survival (%) | 78.9 | 86.6 | 76.5 | NS |

Abbreviations: ADPKD, autosomal dominant polycystic kidney disease; CTIN, chronic tubule-interstitial nephritis; DACD, donor after cardiac death; DN, diabetic nephropathy; eGFR, estimated glomerular filtration rate; GN, glomerulonephritis.
[a]*P*-value <0.05 versus donor score ⩽3 by Scheffé's test.
[b]*P*-value <0.05 versus donor score 4–5 by Scheffé's test.

period. In the CERTPAP study, pathologists' information on their deviation from the mean score value obtained by all participants was associated with only some improvement in κ values for interstitial fibrosis over the course of the study, suggesting that training might allow some improvement in reproducibility. In other areas of pathology, like cancer, reproducibility significantly improved following educational sessions.[23] In the present study, disagreement between the on-call pathologists and the renal pathologist was lower for those with additional training in renal histology. This result suggesting that training may reduce interobserver variability should be interpreted with caution, as in our study only 3 out of the 12 pathologists were considered to have had a special training in renal pathology, and these pathologists only evaluated 20 out of the 122 biopsies. This low number of biopsies precludes evaluation of the predictive value on outcome of additional renal pathology training.

The evaluation of the same paraffin sections by on-call pathologists and the renal pathologist rules out the possibility that different biopsy processing or evaluation of different sections could explain this discrepancy, pointing out that interobserver variability in biopsy scoring constitutes the major reason for the lack of predictive value of donor biopsies evaluated before implantation. This observation raises doubts on the utility of donor biopsies to help in the decision to accept or discard kidneys from ECDs and suggests that, at least in some centers, donor biopsy evaluation may even constitute a barrier to transplant suitable kidneys.[24,25] This point was suggested in the comparison of the discarding rate of retrieved kidneys between the United States and the Eurotransplant group. More than one-third of kidneys from donors aged >60 years and more than half of kidneys procured from donors aged >65 years were discarded in the United States, compared with 8% and 12% of kidneys in the corresponding age groups in the Eurotransplant region. A donor biopsy was more frequently obtained in the United States with the presence of histological damage constituting the main reason to discard retrieved kidneys for transplantation. However, graft survival was not different between the United States and the Eurotransplant group, suggesting that donor biopsy might have constituted an obstacle for transplantation.[24] In Spain, the proportion of retrieved and not implanted kidneys has increased to 25% in 2010.[3] In our study, the overall proportion of kidneys discarded for transplantation was 11.5%, and the proportion of kidneys that were biopsied and discarded owing to histological criteria was 5.5%. This is a relatively low figure and might reflect our strategy to avoid sampling error due to the evaluation of subcapsular samples,[26] biopsies with an insufficient glomerular number,[27] or inadequate interpretation of score discrepancy between kidney pairs from the same donor.[28] Nevertheless, our data do not allow to know whether discarded kidneys would have been suitable for transplantation. In our study, two kidneys discarded for transplantation on histological grounds by the on-call pathologists were considered retrospectively adequate by the renal pathologist.

In the present study, we evaluated chronic damage using a chronic score following the Banff criteria.[29] In other studies, similar chronic scores have been used to discard kidneys for transplantation and to indicate single or dual transplantation.[30,31] Alternatively, other centers have used a simpler approach consisting in the evaluation of the percentage of glomerulosclerosis. This parameter can be

evaluated in frozen and paraffin sections, and it is the most reproducible lesion[32] as confirmed in the present study, but its predictive value on outcome is insufficient.[17,33–35] The preference to use a chronic score instead of the percentage of glomerulosclerosis is based on the observation that evaluation of different histological parameters improves the predictive value of donor biopsy in comparison to the evaluation of one single parameter.[16] Nevertheless, none of these scoring systems has been properly validated, and its superiority over clinical criteria has not been prospectively evaluated. At present, the only well-established strategy to successfully use kidneys from aged donors is shortening cold ischemia time.[36,37]

However, it should be taken into consideration that the retrospective design of our study may be associated with uncontrolled sources of bias. The main limitation is the relative short follow-up, not allowing us to evaluate the predictive value of donor biopsy scores on long-term graft survival. Nevertheless, the lack of association between scores obtained by on-call pathologist and 12-month renal function, or the composite outcome variable, seem sufficient reasons to distrust on the utility of biopsy scoring by multiple observers. Additionally, GFR was estimated but not measured, and it has been shown that prediction equations do not allow a rigorous assessment of renal function in kidney transplant recipients.[38] Finally, on-call pathologists were informed on donor age and gender before biopsy evaluation, but this was not the case in the retrospective evaluation of renal pathologist. Thus, knowledge of donor age and gender may have influenced on-call pathologists scoring.

In conclusion, finding that donor histology and graft outcome were correlated when the biopsy was evaluated by renal pathologists, but not when they were evaluated by on-call pathologists, allows to conclude that a specific training on renal pathology is recommended to optimize the use of kidneys retrieved from ECDs. Biopsy evaluation by non-renal pathologists without specific training may represent an obstacle to transplant a proportion of suitable kidneys or, viceversa, may result in transplantation of kidneys that should be discarded because of too severe histology changes.

## MATERIALS AND METHODS

### Donor biopsies

Since January 2009, all kidneys offered to our renal transplant unit were accepted for clinical, macroscopic, and microscopic examination. Macroscopic evaluation was performed by transplant surgeons who checked symmetry between kidneys and adequate anatomical appearance. A wedge kidney biopsy of approximately 5 mm depth was obtained in ECDs, which were defined as donors accomplishing any of the following criteria: donor age of $\geqslant$55 years, history of hypertension or diabetes mellitus, final pre-procurement serum creatinine $>$ 1.5 mg/dl, or stroke as the cause of death.

Renal biopsies were formalin fixed, rapidly paraffin embedded using microwave technology, stained with hematoxilin—eosin, periodic acid–Schiff, and Masson's trichrome and were assessed by the on-call pathologist. Clinical information available at the time of biopsy scoring was donor age and gender. There were a total of 12 different senior pathologists devoted to a specialized area of pathology different from renal pathology who read the pre-implantation biopsies whenever they were on call. Two out of the 12 non-renal pathologists had additional training in renal pathology as they read renal biopsies in the absence of the renal pathologist. One additional pathologist attended a 2-day intensive course on the evaluation of kidney transplant pathology. The course included training on the evaluation of donor and recipient biopsies and the evaluation of reproducibility. Thus, these three pathologists were considered to have an additional training in the evaluation of renal biopsies. Interstitial fibrosis, tubular atrophy, arteriolar hyalinosis, and intimal vascular thickening were graded from 0 to 3 according to the Banff criteria.[29] Minimum sample size for evaluation was 50 glomeruli between both kidneys from the same donor. Glomerulosclerosis was also evaluated from 0 to 3 according to the percentage of globally sclerosed glomeruli in the whole available cortical tissue (0 = none, 1 = 1–10%, 2 = 11–20%, 3 = $>$20%). A final donor score was calculated as the sum of the individual scores in each renal compartment. Thus, this score ranged from 0 to 15. Kidneys with a global score $\leqslant$7 that did not reach grade 3 score in any compartment were accepted for transplantation. Biopsies with glomerulosclerosis between 20% and 30% were accepted if it was predominantly observed in the subcapsular area and the donor score was $\leqslant$7 ($n$ = 9). In pairs of kidneys from the same donor showing discrepancy between scores, i.e., one kidney was $\leqslant$7 and the other $>$7, the following criteria were applied to reach a decision: (a) if the discrepancy was $\geqslant$3 points, a second biopsy from another area of the kidney with the highest score was obtained and kidneys with a confirmed score $>$7 were discarded and (b) if the discrepancy was $<$3, both kidneys were transplanted provided that macroscopic evaluation did not show significant differences.

Donor biopsies were classified into four categories to facilitate statistical analysis: mild damage (donor score $\leqslant$3), intermediate damage (donor score 4–5), advanced damage (donor score 6–7), and unacceptable ($\geqslant$8) for transplantation.

All biopsies were retrospectively graded by our renal pathologist (MS) to evaluate the reproducibility of the biopsy scoring system. This evaluation was blinded to clinical outcomes of the transplant recipients. A subset of 30 biopsies (the initial 10 biopsies performed in 2009, 2010, and 2011) were evaluated again by the renal pathologist to evaluate intra-observer reproducibility.

### Recipients

Kidneys from standard criteria donors were transplanted in recipients aged $<$55 years while kidneys with a donor biopsy score 6–7 were only transplanted in recipients aged $>$55 years, according to an old-for-old policy.

The maintenance immunosuppression regimen was a calcineurin inhibitor (either tacrolimus or cyclosporine), mycophenolate mofetil, and steroids for all patients. Induction therapy with basiliximab was used in recipients of ECDs, and the introduction of the calcineurin inhibitor was delayed in case of oliguria but never beyond the fifth day post transplantation. Induction with thymoglobulin was used in re-transplants, patients with panel reactive antibodies $>$50%, and recipients of a kidney from a donor after cardiac death.

Delayed graft function was defined as dialysis requirement during the first week after transplant once urinary tract obstruction and/or acute rejection were ruled out. Patients with an acute rise of serum creatinine were biopsied, and histology was evaluated

according to the Banff criteria.[29] In few cases, rejection treatment was initiated without performing a diagnostic biopsy, and these episodes were also considered as rejection episodes if there was a clinical response.

eGFR was calculated by the Modification of Diet in Renal Disease 4 formula.[39]

The main outcome variable was 12-month eGFR. Secondary outcome variables were graft survival, death-censored graft survival, and patient survival as well as a composite outcome variable, including time to death-censored graft failure or time to reach an eGFR $< 30$ ml/min per 1.73 $m^2$ in two consecutive visits to further evaluate the temporal association between renal damage score and graft outcome. Patients never reaching an eGFR $\geqslant 30$ ml/min per 1.73 $m^2$ were uncensored at 1 month. Patients dying with a functioning graft were censored if last eGFR was $\geqslant 30$ ml/min per 1.73 $m^2$. Last follow-up was in April 2013.

### Statistics

Results were expressed as frequencies for categorical variables, median and interquartile range for ordinal variables, and mean ± s.d. for normally distributed continuous variables. Comparison between groups was done by chi-squared for categorical variables, Mann–Whitney *U* test for ordinal variables, and Student's *t*-test or analysis of variance with *post-hoc* Scheffe's test for continuous normally distributed variables. Simple regression analysis was used to study the relationship between continuous variables. Stepwise multiple regression analysis was used to analyze the independent variables associated with 12-month eGFR. For this analysis, categorical variables were transformed into dummy variables.

Weighted κ was used to evaluate the intra-observer and interobserver reproducibility of histological scores obtained by on-call and renal pathologist. This statistic measures inter-rater agreement for categorical items taking into account the agreement occurring by chance. In the case of ordinal variables, weighted κ is used as it takes into consideration the seriousness of disagreement.[40] A κ value $< 0.40$ was considered as poor agreement while a κ value between 0.40 and 0.75 was considered as fair-to-good agreement. Bland–Altman plot was used to evaluate the reproducibility of continuous variables. This plot represents the average measure for both raters in the abscissa against the difference of the measure obtained by both raters in the ordinate. The bias is the mean of the difference variable between raters and the coefficient of repeatability is the 95% limits of agreement (1.96 times the s.d. of the difference variable between raters). To further assess the concordance between donor scores obtained by on-call pathologists and renal pathologist, donor score considered as a continuous variable was used to calculate Lin's concordance correlation coefficient, TDI, and empirical distribution of TDI. TDI was calculated using the package agreement from the statistical analysis software R. We have done calculations for 1 and 2 points differences between donor biopsy scores obtained by on-call pathologists and by the renal pathologist to calculate the coverage probability.[41]

Survival curves were calculated by the Kaplan–Meier method, and comparison between groups was done by the log-rank test. Univariate and multivariate Cox regression analysis were used to estimate risk to reach the outcome variable. All *P*-values were two-tailed, and a *P*-value $< 0.05$ was considered significant.

### DISCLOSURE

All the authors declared no competing interests.

### ACKNOWLEDGMENTS

MAA was supported by a predoctoral Instituto Salud Carlos III grant (FI11-0246) and IBT was supported by a Vall Hebron Research Institute (VHIR) grant. This study was partly supported by a Novartis-Sociedad Española de Trasplante, FIS (PI10-02496) and a REDinREN Instituto Carlos III RD 12/0021/0013 grants. We thank Santiago Perez-Hoyos from the Unitat de Suport Metodologia Investigació Biomèdica Vall Hebron Research Institute for statistical advice and Mrs Cristine Halberstadt for English revision.

### REFERENCES

1. Wolfe RA, Ashby VB, Milford EL *et al.* Comparison of mortality in all patients on dialysis, patients on dialysis awaiting transplantation, and recipients of a first cadaveric transplant. *N Engl J Med* 1999; **341**: 1725–1730.
2. Rabbat CG, Thorpe KE, Russell JD *et al.* Comparison of mortality risk for dialysis patients and cadaveric first renal transplant recipients in Ontario, Canada. *J Am Soc Nephrol* 2000; **11**: 917–922.
3. Organización Nacional de Trasplante (ONT). España 2010. Revised in December 2011. Available on. http://www.ont.es/infesp/Paginas/DatosdeDonacionyTrasplante.aspx.
4. United Network for Organ Sharing (UNOS). United States of America 2010. Revised in December 2011. Available on http://optn.transplant.hrsa.gov/latestData/step2.asp.
5. Sung RS, Guidinger MK, Lake CD *et al.* Impact of the expanded criteria donor allocation system on the use of expanded criteria donor kidneys. *Transplantation* 2005; **79**: 1257–1261.
6. Port FK, Bragg-Gresham JL, Metzger RA *et al.* Donor characteristics associated with reduced graft survival: an approach to expanding the pool of kidney donors. *Transplantation* 2002; **74**: 1281–1286.
7. Ojo AO, Hanson JA, Meier-Kriesche H *et al.* Survival in recipients of marginal cadaveric donor kidneys compared with other recipients and wait-listed transplant candidates. *J Am Soc Nephrol* 2001; **12**: 589–597.
8. Pascual J, Zamora J, Pirsch JD. A systematic review of kidney transplantation from expanded criteria donors. *Am J Kidney Dis* 2008; **52**: 553–586.
9. Cohen B, Smits JM, Haase B *et al.* Expanding the donor pool to increase renal transplantation. *Nephrol Dial Transplant* 2005; **20**: 34–41.
10. Fritsche L, Horstrup J, Budde K *et al.* Old-for-old kidney allocation allows successful expansion of the donor and recipient pool. *Am J Transplant* 2003; **3**: 1434–1439.
11. Nyberg SL, Baskin-Bey ES, Kremers W *et al.* Improving the prediction of donor kidney quality: deceased donor score and resistive indices. *Transplantation* 2005; **80**: 925–929.
12. Nyberg SL, Matas AJ, Kremers WK *et al.* Improved scoring system to assess adult donors for cadaver renal transplantation. *Am J Transplant* 2003; **3**: 715–721.
13. Schold JD, Kaplan B, Baliga RS *et al.* The broad spectrum of quality in deceased donor kidneys. *Am J Transplant* 2005; **5**: 757–765.
14. Navarro MD, Lopez-Andreu M, Rodriguez-Benot A *et al.* Significance of preimplantation analysis of kidney biopsies from expanded criteria donors in long-term outcome. *Transplantation* 2011; **91**: 432–439.
15. 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**: 2325–2334.
16. Lopes JA, Moreso F, Riera L *et al.* Evaluation of pre-implantation kidney biopsies: comparison of Banff criteria to a morphometric approach. *Kidney Int* 2005; **67**: 1595–1600.
17. Gaber LW, Moore LW, Alloway RR *et al.* Glomerulosclerosis as a determinant of posttransplant function of older donor renal allografts. *Transplantation* 1995; **60**: 334–339.
18. Kayler LK, Mohanka R, Basu A *et al.* Correlation of histologic findings on preimplant biopsy with kidney graft survival. *Transpl Int* 2008; **21**: 892–898.
19. 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**: 2316–2324.
20. Furness PN, Taub N. International variation in the interpretation of renal transplant biopsies: report of the CERTPAP Project. *Kidney Int* 2001; **60**: 1998–2012.
21. Colvin RB, Cohen AH, Saiontz C *et al.* Evaluation of pathologic criteria for acute renal allograft rejection: reproducibility, sensitivity, and clinical correlation. *J Am Soc Nephrol* 1997; **8**: 1930–1941.
22. Racusen LC, Solez K, Colvin RB *et al.* The Banff 97 working classification of renal allograft pathology. *Kidney Int* 1999; **55**: 713–723.

23. Warth A, Cortis J, Fink L *et al.* Training increases concordance in classifying pulmonary adenocarcinomas according to the novel IASLC/ATS/ERS classification. *Virchows Arch* 2012; **461**: 185–193.
24. Cecka JM, Cohen B, Rosendale J *et al.* Could more effective use of kidneys recovered from older deceased donors result in more kidney transplants for older patients? *Transplantation* 2006; **81**: 966–970.
25. Rudge CJ. Should retrieved kidneys ever be discarded? *Transplantation* 2006; **81**: 974–975.
26. Muruve NA, Steinbecker KM, Luger AM. Are wedge biopsies of cadaveric kidneys obtained at procurement reliable? *Transplantation* 2000; **69**: 2384–2388.
27. Wang HJ, Kjellstrand CM, Cockfield SM *et al.* On the influence of sample size on the prognostic accuracy and reproducibility of renal transplant biopsy. *Nephrol Dial Transplant* 1998; **13**: 165–172.
28. Foss A, Heldal K, Scott H *et al.* Kidneys from deceased donors more than 75 years perform acceptably after transplantation. *Transplantation* 2009; **87**: 1437–1441.
29. Solez K, Colvin RB, Racusen LC *et al.* Banff 07 classification of renal allograft pathology: updates and future directions. *Am J Transplant* 2008; **8**: 753–760.
30. Remuzzi G, Cravedi P, Perna A *et al.* Long-term outcome of renal transplantation from older donors. *N Engl J Med* 2006; **354**: 343–352.
31. Fernandez-Lorente L, Riera L, Bestard O *et al.* Long-term results of biopsy-guided selection and allocation of kidneys from older donors in older recipients. *Am J Transplant* 2012; **12**: 2781–2788.
32. Singh P, Farber JL, Doria C *et al.* Peritransplant kidney biopsies: comparison of pathologic interpretations and practice patterns of organ procurement organizations. *Clin Transplant* 2012; **26**: E191–E199.
33. Edwards EB, Posner MP, Maluf DG *et al.* Reasons for non-use of recovered kidneys: the effect of donor glomerulosclerosis and creatinine clearance on graft survival. *Transplantation* 2004; **77**: 1411–1415.
34. Bajwa M, Cho YW, Pham PT *et al.* Donor biopsy and kidney transplant outcomes: an analysis using the Organ Procurement and Transplantation Network/United Network for Organ Sharing (OPTN/UNOS) database. *Transplantation* 2007; **84**: 1399–1405.
35. Pokorna E, Vitko S, Chadimova M *et al.* Proportion of glomerulosclerosis in procurement wedge renal biopsy cannot alone discriminate for acceptance of marginal donors. *Transplantation* 2000; **69**: 36–43.
36. Carter JT, Chan S, Roberts JP *et al.* Expanded criteria donor kidney allocation: marked decrease in cold ischemia and delayed graft function at a single center. *Am J Transplant* 2005; **5**: 2745–2753.
37. Kayler LK, Magliocca J, Zendejas I *et al.* Impact of cold ischemia time on graft survival among ECD transplant recipients: a paired kidney analysis. *Am J Transplant* 2011; **11**: 2647–2656.
38. Gaspari F, Ferrari S, Stucchi N *et al.* Performance of different prediction equations for estimating renal function in kidney transplantation. *Am J Transplant* 2004; **4**: 1826–1835.
39. Levey AS, Bosch JP, Lewis JB *et al.* A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of Diet in Renal Disease Study Group. *Ann Intern Med* 1999; **130**: 461–470.
40. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* 1968; **70**: 213–220.
41. Eichinger M, Optazaite DE, Kopp-Schneider A *et al.* Morphologic and functional scoring of cystic fibrosis lung disease using MRI. *Eur J Radiol* 2012; **81**: 1321–1329.