Histopathological analysis of pre-implantation donor kidney biopsy: Redefining the Role in the Process of Graft Assessment.

Chairs: Gianluigi Zaza, Lucrezia Furian

Scientific Committee: Jan Becker, David Cucchiari (Secretary), Aiko de Vries, Albino Eccher, Sandrine Florquin, Jesper Kers, Lorna Marson, Marion Rabant, Michele Rossini

Introduction and rationale for consensus

Preimplantation biopsy provides a window on the state of the renal allograft, and it is a valuable decision-making tool in transplantation (mainly in programs from deceased ECD or high risk recovered donors). However, although the well-reported clinical utility of this procedure, its introduction in the daily clinical practice is still debated and poorly standardized. Currently, there is no consensus about several biopsy-related technical issues and the real impact of the histopathological alterations in kidney compartments as a prognostic factor in graft survival and function is not well defined. A review of the current literature evidence and a discussion on the aforementioned issues may facilitate a consensus on this topic and may have a clear clinical impact in kidney transplantation.

PICOs, references and analysis of the literature, draft statement/recommendation

PICO 1: For the evaluation of chronic lesions in ECD kidneys (P), is the needle core biopsy (I) comparable/inferior/superior to wedge biopsy (C) or punch biopsies (with a skin puncher as in PMID 22492825) in terms of representativity of the entire renal parenchyma (O)?

1. Muruve, Steinbecker, et al. Are wedge biopsies of cadaveric kidneys obtained at procurement reliable? Transplantation 2000 Jun 15;69(11):2384-8

In this study, the authors compared the histology obtained from wedge biopsies in 9 donors to the histology obtained from nephrectomy of the same kidneys. Wedge biopsies of donor kidneys overestimated the total amount of glomerulosclerosis, apparently because of a predominance of sclerosis in the kidney's subcapsular region, the area predominantly sampled by the usual wedge biopsy.

2. Yushkov, Dikman, et al. Optimized technique in needle biopsy protocol shown to be of greater sensitivity and accuracy compared to wedge biopsy. Transplant Proc 2010 Sep;42 (7):2493-7

In this study, the authors compared in the same kidney, wedge biopsies to optimized needle biopsies (2 cores of 14g) in terms of GS, IF/TA, cv, ATN, in a cohort of 226 kidneys. ONBT detected more tubular interstitial scarring and arterial intimal fibrous narrowing than WB (P = .00). Difference in glomerular yield or ATN between WB and ONBT was not statistically significant.

The data suggest that there were no statistical differences in sample reliability between ONBT and WB. However, ONBT was found to be significantly more sensitive in identifying allograft tubular interstitial scarring as well as intimal fibrous narrowing. Overall this study provides proof that ONBT is a more reliable and accurate method compared to WB in identifying important parameters of renal allograft.

3. Yong, Kipgen, et al. Wedge Versus Core Biopsy at Time Zero: Which Provides Better Predictive Value for Delayed Graft Function With the Remuzzi Histological Scoring System? Transplant Proc 2015 Jul-Aug;47(6):1605-9.

In this study, the authors compared the predictive value for DGF of 37 wedge biopsies and 30 needle core biopsies, based on GS, IFTA and ah, Remuzzi scoring. The results showed more glomeruli in wedge biopsies and more arterioles in needle. Wedge biopsies were more likely to identify pathology with more glomerulosclerosis, tubular atrophy ($P < .01$), and interstitial fibrosis ($P < .01$). The sensitivity and positive predictive value of Remuzzi ≥ 4 for predicting DGF was better on wedge biopsy (45.5% versus 0%; $P < .01$ and 62.5% versus 0%; $P < .01$, respectively). Wedge biopsies were safe and superior to core biopsies for identifying clinically significant histopathological findings on pre-implantation renal biopsy.

4. Haas M, Segev D et al. Arteriosclerosis in kidney from healthy live donors. Comparison of wedge and needle core perioperative biopsies. Arch Pathol Lab Med, Vol 132, 2008

In this study, the authors compared in the same kidney wedge and needle core biopsies in a cohort of 36 living donors. Wedge biopsies retrieved more glomeruli, the same number of interlobular arteries but fewer arcuate arteries. Arteriosclerosis was more severe in needle biopsies, but there was no difference regarding GS, IFTA and ah. Needle core biopsies are superior to wedge for evaluating vascular changes in donor kidneys.

5. Mazzuco, Magnani et al. The reliability of donor renal biopsies in predicting the kidney state. A comparative single center study on 154 untransplanted kidneys . NDT 2010, 25: 3401

In this study, the authors compared the result of 118 needle core biopsies and 36 wedge biopsies to histology from the nephrectomy in the same kidneys, with regards to IFTA, vascular damage and GS. Agreement between biopsies and kidney were globally similar with high score for vascular damages, and a globally better score for needle biopsies. Global GS score requires some caution.

6. Bago Horwath S, Kozakowski N et al. The cutting (w) edge- comparative evaluation of renal baseline biopsies obtained by 2 different methods. NDT 2012 27: 3241

In this study, the authors compared 147 punch biopsies and 114 WG with respect to number of glomeruli and arterial vessels, GS, IFTA and cv. The results were compared to post transplant biopsies performed for cause within 2 months. Wedge contained more glomeruli but less arteries. Punch biopsies were superior to diagnose IFTA and cv. The use of skin B tools is safe and effective.

7. Husain S.A, Shah V et al. Impact of deceased donor kidney procurement biopsy technique on histologic accuracy. KIR 2020 (5), 1908.

In this study the authors compared 171 procurement wedge biopsies and 221 procurement needle biopsies to reperfusion biopsies in the same kidneys, using a validated score based on GS>11% , IFTA >or=2 and moderate/severe vascular disease. Biopsies were mainly interpreted by non-renal pathologists. Wedge biopsies found more glomerulosclerosis and IFTA and needle found more vascular disease. Biopsy technique and pathologist training were not associated with histologic accuracy, but a larger number of glomeruli was associated with a better concordance.

8. Liapis H, Gaut JP et al. Banff histopathological consensus criteria for preimplantation biopsies. American Journal of Transplantation 2017; 17: 140–150. Banff working group

This working group concludes that this study shows some superiority of frozen wedge versus needle core biopsies. This may be related to the lack of standardization of needle biopsies. Different needle gauges lead to different degrees of sampling that are not apparent in standardized wedge biopsies. Banff participants expressed the opinion that there are technical difficulties in cutting 18-gauge needle core biopsies, resulting in inadequate material for frozen section interpretation. Needle biopsies are perceived to be more likely to damage larger caliber medullary vessels. The possibility of using punch biopsies instead was suggested, but most in the working group thought that an ample wedge biopsy with good cortical sampling is more likely to yield sufficient number of glomeruli and is likely superior to needle or punch biopsies. They also found that wedge biopsies were often subcapsular and not deep enough to capture vessels . In conclusion, the working group found wedge biopsies to be superior to needle core biopsies.

ANALYSIS OF THE LITERATURE

A majority of centers use wedge biopsies for obtaining zero-time kidney transplant histology, whereas other centers use needle biopsies. To date, there is no consensus with regards to the better technique. Many studies have shown that wedge biopsies (WB) provide more glomeruli but may overestimate glomerulosclerosis (Muruve et al; Yong et al, Husain et al) because of their more superficial nature. On the other hand, WB have been shown to underestimate the extent of arterial intimal thickening (Haas et al). Another study (Yong et al, 67 biopsies) showed that wedge biopsy had far more glomeruli, more IFTA and more glomerulosclerosis and were superior to needle biopsy for predicting DGF. However, in that study, wedge and needle were not compared in the same patients and the endpoint is DGF which can be multifactorial.

Only 2 studies have directly compared WB and NB in the same kidney (Yushkov et al, Haas et al). In 226 donors, Yushkov et al. found that optimized needle biopsies were significantly more sensitive in identifying allograft tubular interstitial scarring as well as intimal fibrous narrowing than WB. However, the technique of NB implied 2 cores of 14-gauge needles. Haas et al. also found more severe arteriosclerosis in NB, partly due to the higher number of arcuate arteries in NB compared to WB, but this study was performed in healthy living donors. Several biopsies have compared the results of donor biopsies to the histology of the nephrectomy of the same kidney (Muruve et al., n=9; Mazzuco et al.; n=154) or to early post-transplant biopsies (Bago et al, n=271; Husain et al; n=392). Muruve et al. found that WG overestimate glomerulosclerosis whereas Mazzuco et al. found in 118 NB and 36 WB that agreement between biopsies and kidney were globally similar with both

techniques with high score for vascular damages, and a globally better score for needle biopsies. Husain et al. found that in 171 WB and 221 NB, WB found more glomerulosclerosis and IFTA and NB found more vascular disease, but there was no difference in the concordance with reperfusion biopsies. Only a larger number of glomeruli was associated with a better concordance.
Only one study (Bago Horwath et al.) used punch biopsies (PB) for preimplantation biopsies and compared the results of 147 PB and 114 WG to post-transplant biopsies performed for cause within 2 months. WB contained more glomeruli but less arteries but PB were superior to diagnose IFTA and cv.

Statement: For the evaluation of chronic lesions in ECD kidneys (P), needle core biopsy and wedge or punch biopsy are suitable even though differences may be found in terms of glomerular and vascular assessment.

NB: wedge biopsies may overestimate GS and contain less arteries, whereas needle biopsies may contain less glomeruli, but more arteries.
Recommendations regarding the methods of wedge biopsies (Randhawa et al.) and needle biopsies have been proposed (Yushkov et al)
Punch biopsies have potentially similar suitability, although more evidence is required to include it in the statement.

Quality of evidence: moderate

Strength of recommendation: strong

PICO N 2: For the evaluation of chronic lesions in ECD kidneys (P), is the frozen section (I) comparable/inferior/superior to paraffin embedded section (C) in terms of reliability of the reading from pathologists?

REFERENCES

9. Carpenter, Husain, et al. Procurement Biopsies in the Evaluation of Deceased Donor Kidneys.
10. Sagasta, Sanchez-Escuredo, et al. 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.
11. Teixeira, Freire de Carvalho, et al. Evaluation of Frozen and Paraffin Sections Using the Maryland Aggregate Pathology Index Score in Donor Kidney Biopsy Specimens of a Brazilian Cohort.
12. Goumenos, Kalliakmani, et al. The prognostic value of frozen section preimplantation graft biopsy in the outcome of renal transplantation.

ANALYSIS OF THE LITERATURE

One large study included kidneys where more than one biopsy was taken (Carpenter et al, n=116). They found that procurement biopsies are poorly reproducible and do not correlate well with paraffin- embedded reperfusion biopsies, especially in the degree of glomerulosclerosis. For kidneys on which more than one procurement biopsy was performed (n=116), category agreement was found in only 64% of cases (k=0.14). For all kidneys (n=270), correlation between procurement and reperfusion biopsies was poor. This discrepancy was most pronounced when categorizing percentage of glomerulosclerosis, which had 63% agreement (k=0.15). Interstitial fibrosis/tubular atrophy and vascular disease had agreement rates of 82% (k=0.13) and 80% (k=0.15), respectively.

A smaller study (Sagasta et al, n=92) found that agreement between observers (on call pathologist versus trained pathologist) using the same frozen section was weaker than the correlation between frozen section and paraffin-embedded section. 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). This comparison revealed that the trained pathologist assigned higher scores than the on-call pathologists and, accordingly, renal transplant acceptance was higher by the on-call pathologists than by the trained pathologist. Furthermore, the trained pathologist assigned higher scores when using frozen than with permanent sections and, accordingly hypothetical organ acceptance comparing both techniques excluded more organs from transplantation when using frozen than permanent sections.

Another smaller study (Goumenos et al, n=74) showed that frozen section and paraffin section showed comparable histological changes. Although glomerulosclerosis and arteriolosclerosis were underestimated, whereas acute tubular necrosis and interstitial

fibrosis were overestimated, in the frozen sections compared to permanent ones, but those differences were not statistically significant.

Teixera et al (n= 262) used an aggregate score (MAPI) to assess agreement between frozen section and paraffin-embedded biopsies, showing improved Kappa coefficient when the score was used than the individual parameters. They performed a retrospective review of pathological reports of frozen sections (on-call pathologist) and their corresponding permanent sections (trained pathologist). Kappa values ranging from 0.29 to 0.51 were obtained when MAPI parameters were separately evaluated. When the score was used, the coefficient was 0.59.

RECOMMENDATION/STATEMENT: for the evaluation of chronic lesions in ECD kidneys the frozen section is inferior to paraffin embedded section in terms of reliability of the reading from pathologists.

RECOMMENDATION/STATEMENT: Frozen sections should not be considered as a first intention option; however, it could be suitable for use in highly selected cases like particular urgency or specific contexts.

Quality of evidence: High

Strength of recommendation: Weak against

PICO N 3: For score assessment of pre-implantation kidney biopsy in the evaluation of ECD (P) is the experienced renal pathologist (I) comparable/inferior/superior to on-call pathologist (C) in terms of reproducibility and accuracy of the histological report (O)?

REFERENCES

1. Sagasta, Sanchez-Escuredo. 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.
2. Antonieta Azancot, Moreso, et al. The reproducibility and predictive value on outcome of renal biopsies from expanded criteria donors.
3. Girolami I, Gambaro G, Ghimenton C et al . Pre-implantation kidney biopsy: value of the expertise in determining histological score and comparison with the whole organ on a series of discarded kidneys. J Nephrol. 2020 Feb;33(1):167-176. doi: 10.1007/s40620-019-00638-7. Epub 2019 Aug 30.PMID: 3147181

ANALYSIS OF THE LITERATURE

In a study of 92 biopsies, 78 from transplanted and 14 from non-transplanted kidneys, the correlation between the on-call pathologists and the trained pathologist was weak in all the parameters on frozen section (Sagasta et al). The trained pathologists tended to assign Pre-implantation Biopsies from Expanded Criteria Kidney Donors higher Remuzzi scores than the on-call pathologists.

A larger study by Antonieta Azancot et al. on 127 kidneys found agreement between the scores of general and specialist pathologist was poor or fair for all variables other than GS which was near perfect. General pathologists tended to have higher aggregate scores thus a tendency to overcall chronic damage and there was no association between readings of on-call pathologists and outcome, whilst evaluation of biopsies by a renal pathologist was significantly and independently associated with estimated 12-month glomerular filtration rate and composite graft outcome.

In the paper of Girolami et al. 46 discarded kidneys were identified with their corresponding biopsies. The biopsies were reviewed by three general and two specialist pathologists, blinded to the original report, according to Remuzzi score. The intraclass correlation coefficient (ICC) was calculated for both groups. Specialist pathologists achieved higher values of ICC, reaching excellent or good agreement in most of the parameters, while general pathologists' values were mainly fair or good. As a quote of transplantable kidneys may be erroneously discarded due to lack of specialty, the availability of an expert reading pathologist would improve organ utilization.

Notably in the Banff Histopathological Consensus Criteria for Preimplantation Biopsies (Liapis et al. 2017), although they did not compare the interpretation by renal pathologists versus general pathologists, they recommend a training of pathologists assigned to donor biopsy interpretation.

RECOMMENDATION/STATEMENT: For score assessment of pre-implantation kidney biopsy in the evaluation of ECD the experienced renal pathologist is superior to non-experienced pathologist in terms of reproducibility and accuracy of the histological report

Quality of evidence: High (Level 2)

Strength of recommendation: Strong for

PICO 4: In the quantification of the chronic damage in ECD kidneys (P), is glomerulosclerosis (I) more reproducible (O) in comparison with other parameters (interstitial fibrosis, tubular atrophy, wall/lumen ratio, arteriolar hyalinosis) (C)?

PAPERS

4. Snoeijs MG J et al . Histological assessment of pretransplant kidney biopsies is reproducible and representative. Histopathology 2010, 56, 198-202
5. Girolami I, Gambaro G, Ghimenton C et al . Pre-implantation kidney biopsy: value of the expertise in determining histological score and comparison with the whole organ on a series of discarded kidneys. J Nephrol. 2020 Feb;33(1):167-176. doi: 10.1007/s40620-019-00638-7. Epub 2019 Aug 30.PMID: 3147181
6. Marsh Jon et al. Development and Validation of a Deep Learning Model to Quantify Glomerulosclerosis in Kidney Biopsy Specimens JAMA Netw Open. 2021 Jan; 4(1): e2030939
7. Husain Ali et al. Reproducibility of Deceased Donor Kidney Procurement Biopsies cJASN 15: 257–264, 2020
8. Verduzco, Batal, Mohan and Husain Reproducibility of Chronic Changes on High-Quality Deceased Donor Kidney Allograft Biopsies, KIR (2022) 7, 899-891

SUMMARY

In a study of 44 donor biopsies (50% needle, 50% wedge), the glomerulosclerosis, cv, TA, IF were scored by 3 independent pathologists. The ICCs was highest for glomerulosclerosis (0.87) than cv (0.51), TA (0.71) and IF (0.35). ICC were similar for wedge and needle biopsies (Snoeijs et al.)

In a more recent study (Girolami I et al), 46 discarded kidneys were identified with their 75 corresponding biopsies (83 % wedge and 17% needle). Biopsies were reviewed by 3 general and 2 specialist pathologists. Specialist pathologists achieved higher values of ICC with excellent or good agreement, while general pathologists' values were mainly fair or good. Interestingly, the ICC was highest for GS and was comparable between general and specialists, whereas ICC for IFTA and vascular changes was poor to fair in general pathologists and good to excellent in specialists, fair for IFTA. The percent GS was however significantly higher in the biopsies than in discarded organs, demonstrating a "true" sampling error of GS as the majority of biopsies were wedge biopsies.

Using artificial intelligence, a deep neural network segmented normal and globally sclerotic glomeruli in 98 HES frozen and 51 FFPE whole-slide images from 83 donor kidneys biopsies, to quantify percent global glomerulosclerosis. Annotation by 3 expert pathologists served as ground truth. A total of 1544 globally sclerosed and 6914 nonglobally sclerosed were labelled in 149 images. The study demonstrated higher

performance of the DL model than pathologists. Model accuracy further increased by pooling multiple sections, resulting in decreased likelihood of organ discard by 37%. However, this study did not compare the reproducibility of GS to other chronic parameters in the biopsy.

Two studies form the same center in Columbia university focused on the reproducibility of chronic scores in sequential biopsies from the same donor:
Husain et al. included 1010 procurement biopsies among which 606 had more than one procurement biopsies. Information about GS, IF, TA, cv were retrieved from the reports. A score from 0 to 3 was assigned for each parameter. The agreement between sequential biopsies reports for kidney that underwent multiple procurement biopsies was evaluated. There was poor overall agreement for the 3 histologic compartments, and agreement was highest for vascular disease and lowest for GS.
More recently, they compared protocol kidney biopsies performed at day 7 and 14 in 69 patients and obtained the reported GS, IFTA, cv, ah scores. Agreement between day 7 and day 14 was best for arteriosclerosis (concordance 78%, k=0.60). For GS, only a moderate correlation between both time points was found (r^2=0.25).

In conclusion, GS seems to be more reproducible between pathologists looking at the same biopsy, compared to IF/TA or vascular changes. However, GS in sequential biopsies from the same donor is not the most reproducible chronic feature.

RECOMMENDATION:

In the quantification of the chronic damage in ECD kidneys, glomerulosclerosis is more reproducible in comparison with other parameters (interstitial fibrosis, tubular atrophy, wall/lumen ratio, arteriolar hyalinosis.

Quality of evidence: Weak

Strength of recommendation: slight in favor

PICO N 5: In the quantification of the chronic damage in ECD kidneys (P) is measurement of histological variables with digital pathology (I) comparable/inferior/superior (O) if compared with light microscopy (C)?

REFERENCES

1. Girolami I, Pantanowitz L, Marletta S, Hermsen M, van der Laak J, Munari E, et al. Artificial intelligence applications for pre-implantation kidney biopsy pathology practice: a systematic review.
   Altini N, Cascarano GD, Brunetti A, Marino F, Rocchetti MT, Matino S, et al. Semantic Segmentation Framework for Glomeruli Detection and Classification in Kidney Histological Sections.
   Bevilacqua V, Pietroleonardo N, Triggiani V, Brunetti A, Di Palma AM, Rossini M, et al. An innovative neural network framework to classify blood vessels and tubules based on Haralick features evaluated in histological images of kidney biopsy.
2. Luo Y, Liang J, Hu X, Tang Z, Zhang J, Han L, et al. Deep Learning Algorithms for the Prediction of Posttransplant Renal Function in Deceased-Donor Kidney Recipients: A Preliminary Study Based on Pretransplant Biopsy.
3. So WZ, Teo RZC, Ooi LY, Goh BYS, Lu J, Vathsala A, et al. Multi-photon microscopy for the evaluation of interstitial fibrosis in extended criteria donor kidneys: A proof-of-concept study.
4. Cascarano GD, Debitonto FS, Lemma R, Brunetti A, Buongiorno D, De Feudis I, et al. A neural network for glomerulus classification based on histological images of kidney biopsy. BMC medical informatics and decision making.
5. Marsh JN, Matlock MK, Kudose S, Liu TC, Stappenbeck TS, Gaut JP, et al. Deep learning global glomerulosclerosis in transplant kidney frozen sections. IEEE Transactions on Medical Imaging.
6. Yi Z, Salem F, Menon MC, Keung K, Xi C, Hultin S, et al. Deep learning identified pathological abnormalities predictive of graft loss in kidney transplant biopsies.
   Salvi M, Mogetta A, Meiburger KM, Gambella A, Molinaro L, Barreca A, et al. Karpinski Score under Digital Investigation: A Fully Automated Segmentation Algorithm to Identify Vascular and Stromal Injury of Donors' Kidneys.
7. Salvi M, Mogetta A, Gambella A et al (2021) Automated assessment of glomerulosclerosis and tubular atrophy using deep learning. Comput Med Imaging Graph Off J Comput Med Imaging Soc 90:101930
8. Eccher A, Neil D, Ciangherotti A, Cima L, Boschiero L, Martignoni G, Ghimenton C, Chilosi M, Giobelli L, Zampicinini L, Casartelli M, Brunelli M.Hum Pathol. 2016 Jan;47(1):115-20. doi: 10.1016/j.humpath.2015.09.012. Epub 2015 Sep 30.PMID: 26547252

ANALYSIS OF THE LITERATURE

The study by Altini et al (n=26, 2500 glomeruli) aimed to detect and classify glomeruli in kidney biopsies using a two models segmentation convolutional neural network. Global accuracy was higher than 0.98 with precision in classifying healthy and sclerosed glomeruli ranging 0.834–0.935 and 0.806–0.976.

The paper of Bevilacqua et al tested a Computer Aided Diagnosis system for segmentation and discrimination of blood vessels versus tubules from 10 biopsies in

the kidney tissue through the elaboration of histological images: regions Of Interest identified were in total 221:71 vessels and 150 tubules. Results determined that the supervised artificial Neural Network approach was consistent and reveals good performance, after a training phase based on vessels and tubules samples. Accuracy was higher than 0.93 with precision higher than 0.88 in validation set and higher than 0.91 in test set.

Luo et al used donor kidney biopsy WSIs as a feature in addition to clinical characteristics for graft function prediction, building neural network models to predict stable eGFR and RGF in donor kidney recipients underwent pretransplantation biopsy. They tested six prediction models on 219 WSI. Overall, donor kidney biopsy WSIs were a useful predictor for graft function recovery, showing distinct improvements in the prediction performance of the deep learning algorithm plus the clinical characteristics model. Compared with the clinical data model, the area under the receiver operating characteristic (ROC) curve (AUC) of the clinical data plus image model for eGFR classification increased from 0.69 to 0.83. In addition, the predictive performance for RGF increased from 0.66 to 0.80.

In a proof-of-concept study, So et al reported noteworthy differences in Multiphoton Microscopy derived collagen parameters between donor kidneys of varying KDPI scores. They evaluated the amount (CART) and quality (CRI) of collagen deposition in 20 preimplantation biopsies. Although CAR values were identical across all samples, biopsies classified with > 85% KDPI demonstrated a significantly higher CART (51.94 vs. 45.61; p = .011) compared to biopsies with 20–85% KDPI percentages. Conversely, they had lower CRI compared to biopsies with 20–85% KDPI scores (4.15 vs. 4.53; p = .025).

Cascarano et al collected 26 digital slides taken from the kidneys of 19 donors with Periodic Acid-Schiff staining with the aim to develop a neural network able to detect and classify glomeruli. The workflow allowed the classification of sclerotic and non-sclerotic glomeruli with good performances: 0.99 accuracy, 1.00 precision.

Marsh et al developed a deep-learning model for glomerulosclerosis on a population of mixed wedge and core kidney biopsy cases: 98 frozen and 51 permanent sections. Glomeruli counts were compared against annotation ground truth, with accuracy assessed by Pearson correlation coefficient. The model correlated very well with the pathologists' annotations, with a correlation coefficient higher than 0.900.

Salvi et al developed two models: RENFAST (Rapid EvaluatioN of Fibrosis And vesselS Thickness) for vessels and interstitial fibrosis detection and RENTAG (Robust EvaluatioN of Tubular Atrophy & Glomerulosclerosis) for glomeruli and tubuli detection and classification. The RENFAST algorithm is developed and tested on 350 periodic acid–Schiff images for blood vessel segmentation and on 300 Massone's trichrome stained images for the detection of renal fibrosis. In the test set, the algorithm exhibits excellent segmentation performance in both blood vessels (accuracy: 0.8936) and fibrosis (accuracy: 0.9227). The algorithm takes an average computational time 2.91 s against 20 min for pathologist assessment. RENTAG was developed using 61 WSIs

for glomerulosclerosis assessment while 22 WSIs were employed for tubular atrophy quantification. Algorithm showed a Dice score of 0.95 and 0.91 for glomeruli and CNN model tubuli detection with 100% sensitivity and PPV and little time of computation required.

Eccher et al evaluated 62 consecutive, previously reported preimplantation kidney biopsies scanned with the ScanScope Digital Slide Scanner. The slides were assessed for percent glomerulosclerosis, tubular atrophy, interstitial fibrosis and vascular narrowing using the Remuzzi criteria by two pathologists, one using glass slides and the other using the whole-slide images viewed on a widescreen computer monitor. After a 2-week washout period, all of the slides were re-assessed by the same pathologists using the opposite mode of reporting to that used in the first evaluation. Very high glass-digital intra-observer concordance was achieved for the overall score and for individual grades by both pathologists (κ range, 0.841-0.973).

RECOMMENDATION: in the quantification of the chronic damage in ECD kidneys measurement of histological variables with digital pathology is potentially comparable with light microscopy

Quality of evidence: High

Strength of recommendation: Strong for

Artificial intelligence can help pathologist in assessment of histological variables in kidney also reducing interobserver variability. The future potential in terms of 1) infrastructure and organization of care and 2) algorithmic assessment of digital pathology and AI need further evidence.

PICO 6 In the quantification of the chronic damage in ECD kidneys (P) is measurement of histological variables with the aid of special stainings (Periodic-Acid Schiff, Silver, Picro Sirius Red, Trichrome stainings) (I) comparable/inferior/superior (O) if compared with Haematoxylin & Eosin alone (C)?

REFERENCES

No suitable references were identified.

ANALYSIS OF THE LITERATURE

The literature search did not identify articles that fit the search criteria related to the PICO question. In general, the Scientific Committee strongly beliefs that for any renal pathology setting, only performing an H&E staining is inferior to a dedicated panel of special histochemical stainings that also includes Periodic-acid Schiff, Silver, Trichrome and/or Picro Sirius Red stainings. However, in the setting of (on-call) organ utilization decision making specifically, where optimal decision making competes with time constrains, processing of special histochemical stains (either performed on frozen sections or fast formalin-fixation protocols) will likely result in an unwanted delay of the organ transplant procedure and accompanied increase in ischemia time for several hours.

STATEMENT PROPOSAL:

In the quantification of chronic damage in ECD kidneys, the use of additional histochemical stainings (including, but limited to PAS, Silver, Trichrome and/or Picro Sirius Red) is superior to the use of H&E alone in any diagnostic kidney pathology context but can likely not be performed under time constraints in the context of (on-call) organ utilization decision making.

Quality of evidence: Low, Expert opinion (CEBM Level 5, GRADE Level D)

Strength of recommendation: Strong for

PICO 8 In the quantification of the chronic damage in ECD kidneys (P), is glomerulosclerosis percentage (I) more representative than other parameters (interstitial fibrosis, tubular atrophy, arteriolar hyalinosis and cv score) (C) to predict the graft survival, graft function, primary non-function (O)?

REFERENCES

1. Stewart DE, Foutz J, Kamal L, Weiss S, McGehee HS, Cooper M, et al. The Independent Effects of Procurement Biopsy Findings on Ten-Year Outcomes of Extended Criteria Donor Kidney Transplants. Kidney International Reports. 2022 May 30;7(8):1850-1865. PMCID: PMC9366372

2. Anglicheau D, Loupy A, Lefaucheur C, Pessione F, Létourneau I, Côté I, K et al. A Simple Clinico-Histopathological Composite Scoring System Is Highly Predictive of Graft Outcomes in Marginal Donors. Am J Transplant. 2008 Nov;8(11):2325-34. PMID: 18785957

3. Cheungpasitporn W, Thongprayoon C, Vaitla PK, Chewcharat A, Hansrivijit P, Koller FL, et al. Degree of Glomerulosclerosis in Procurement Kidney Biopsies from Marginal Donor Kidneys and Their Implications in Predicting Graft Outcomes. J Clin Med. 2020;9(5):1469. PubMed PMID: 32422905.

4. Bodzin AS, Leiby BE, Ramirez CB, Frank AM, Doria C. Expanded criteria donor kidneys where the paired kidney is discarded owing to biopsy results: A concept that needs revision. Experimental and Clinical Transplantation. 2014 01 Dec;12(6):499-505. PubMed PMID: 600749956.

5. Kayler LK, Mohanka R, Basu A, Shapiro R, Randhawa PS. Correlation of histologic findings on preimplant biopsy with kidney graft survival. Transplant International. 2008 September;21(9):892-8. PubMed PMID: 352175703.

6. Sung RS, Christensen LL, Leichtman AB, Greenstein SM, Distant DA, Wynn JJ, et al. Determinants of discard of expanded criteria donor kidneys: impact of biopsy and machine perfusion. American Journal of Transplantation. 2008;8(4):783-92. PubMed PMID: 18294347.

ANALYSIS OF THE LITERATURE

In a recent study [1], Stewart et al., analyzing a large dataset of 3851 extended criteria donors (ECDs) recovered in the United States from 2008 to 2012, reported a significant effect of glomerulosclerosis (GS>10%) on kidney graft survival, even after adjustment for potentially confounding donor and recipient variables. Whilst the effects of interstitial fibrosis and vascular changes were attenuated after adjustment. The BARETO (Biopsy, Anatomy, and Resistance Effects of Transplant Outcomes) study found a clinically and statistically significant effect of GS on 10-year graft survival among ECD kidney transplants. Kidneys having GS>10% were found to have 18% higher risk of graft failure compared with kidneys with GS 0-5%.

The effect waned beyond 10%, suggesting little or no incremental risk associated with a GS of 20% compared with a GS of 10%. Regarding vascular changes, their data suggest a possible meaningfully

large effect of mild-moderate (>25%) or worse vascular changes on long-term graft survival. Interstitial fibrosis seemed to have minimal, if any, prognostic value. These results were in line with those previously published by Anglicheau et al [2] demonstrating that GS was an independent histological predictor of low eGFR at 1 year and death-censored graft survival. Also in this case, the cut-off of GS more that 10% was the most significant.

Cheungpasitporn et al. [3] analyzed kidney graft outcomes related to the degree of GS in numerous datasets (>22000 kidneys) ECDs with a kidney donor profile index (KDPI) score > 85% from 2005 to 2014 [4]. They found that GS >10% is independently related to increased risk of graft loss. Kidneys with >10% GS were associated with 27% higher risk of graft failure compared to kidneys with 0-10%. Of note, there was no difference in graft survival between 11-20% and >20% GS.

These results were in contrast with those previously published by Bodzin et al. [4] using the Organ Procurement and Transplant Network (OPTN) data. Multivariate analysis demonstrated that kidneys from ECDs with 0-5% GS had no significant differences in graft function compared with those having more than 10% GS.

Additionally, Kayler et al. [5], analyzing a large dataset of kidney transplant recipients (n:597) showed that only the presence of moderate arteriosclerosis and/or moderate arteriolosclerosis (MA), defined as >or=25% luminal compromise, was a significant predictor of graft outcome in standard criteria donors (multivariate P=0.01) and in ECDs as defined by UNOS criteria (univariate P=0.02). One-, 3-, and 5-year overall allograft survival MA was 71%, 58%, and 40%, respectively. Increasing degree of GS was not associated with earlier graft failure on multivariate analysis (P=0.36).

GS>20% and interstitial fibrosis>25% had a low frequency in the material reviewed, likely reflecting their organ use practices and did not have a demonstrable effect on graft outcome.

Finally, Sung et al. [6], in another large multivariate analysis performed using the Scientific Registry of Transplant Recipients (SRTR)/Organ Procurement and Transplantation Network (OPTN) data, found that GS was not reliably associated with DGF or graft failure.

Therefore, based on the aforementioned literature evidence, we cannot draw definitive conclusions regarding the impact of the GS on kidney transplant clinical outcomes. We also cannot consider this histological feature more representative than other parameters (interstitial fibrosis, tubular atrophy, arteriolar narrowing, arteriosclerosis) to predict graft function and survival. No studies have considered the primary non function as the target clinical outcome. Further studies must address these research gaps.

STATEMENT PROPOSAL:

The degree of GS in procurement kidney biopsies from ECDs is associated with an increased risk of long-term graft failure. However, GS should not be used in isolation from other biopsy findings for individualized organ acceptance decision.

Quality of evidence: Moderate (Level 3)

Strength of recommendation: Strong for