

# Automatic Measurement of Kidney and Liver Volumes from MR Images of Patients Affected by Autosomal Dominant Polycystic Kidney Disease

Maatje D.A. van Gastel,<sup>1,2</sup> Marie E. Edwards,<sup>1</sup> Vicente E. Torres,<sup>1</sup> Bradley J. Erickson,<sup>3</sup> Ron T. Gansevoort,<sup>2</sup> and Timothy L. Kline<sup>1,3</sup>

<sup>1</sup>Division of Nephrology and Hypertension, Department of Internal Medicine; <sup>2</sup>Department of Nephrology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands and <sup>3</sup>Department of Radiology, Mayo Clinic, Rochester, Minnesota

## ABSTRACT

**Background** The formation and growth of cysts in kidneys, and often liver, in autosomal dominant polycystic kidney disease (ADPKD) cause progressive increases in total kidney volume (TKV) and liver volume (TLV). Laborious and time-consuming manual tracing of kidneys and liver is the current gold standard. We developed a fully automated segmentation method for TKV and TLV measurement that uses a deep learning network optimized to perform semantic segmentation of kidneys and liver.

**Methods** We used 80% of a set of 440 abdominal magnetic resonance images (T2-weighted HASTE coronal sequences) from patients with ADPKD to train the network and the remaining 20% for validation. Both kidneys and liver were also segmented manually. To evaluate the method's performance, we used an additional test set of images from 100 patients, 45 of whom were also involved in longitudinal analyses.

**Results** TKV and TLV measured by the automated approach correlated highly with manually traced TKV and TLV (intraclass correlation coefficients, 0.998 and 0.996, respectively), with low bias and high precision ( $<0.1\% \pm 2.7\%$  for TKV and  $-1.6\% \pm 3.1\%$  for TLV); this was comparable with inter-reader variability of manual tracing ( $<0.1\% \pm 3.5\%$  for TKV and  $-1.5\% \pm 4.8\%$  for TLV). For longitudinal analysis, bias and precision were  $<0.1\% \pm 3.2\%$  for TKV and  $1.4\% \pm 2.9\%$  for TLV growth.

**Conclusions** These findings demonstrate a fully automated segmentation method that measures TKV, TLV, and changes in these parameters as accurately as manual tracing. This technique may facilitate future studies in which automated and reproducible TKV and TLV measurements are needed to assess disease severity, disease progression, and treatment response.

JASN 30: 1514–1522, 2019. doi: <https://doi.org/10.1681/ASN.2018090902>

Total kidney volume (TKV) is a prognostic enrichment biomarker in patients with polycystic kidney disease (PKD) that predicts future renal function decline.<sup>1–4</sup> In early disease, it has additional value over renal function measurements, which can stay within normal ranges for a prolonged period of time due to hyperfiltration of remaining nephrons.<sup>3</sup> As such, TKV is recognized by the Food and Drug Administration and the European Medicines Agency as a prognostic enrichment biomarker for disease progression in autosomal dominant polycystic kidney disease (ADPKD).<sup>5,6</sup> In addition, TKV is used in clinical

care to assess the risk of individual disease progression per patient and select patients with rapid disease progression for tolvaptan treatment or clinical trials. Moreover, in clinical trials, TKV

Received September 7, 2018. Accepted April 10, 2019.

Published online ahead of print. Publication date available at [www.jasn.org](http://www.jasn.org).

**Correspondence:** Dr. Timothy L. Kline, Department of Radiology, Mayo Clinic, 200 First Street SW, Rochester, MN 55905. Email: [kline.timothy@mayo.edu](mailto:kline.timothy@mayo.edu)

Copyright © 2019 by the American Society of Nephrology

is often used as a primary or secondary end point to assess treatment effects.

In patients with ADPKD, liver cyst development has been reported in >94% of patients.<sup>7</sup> Having a polycystic liver is negatively associated with quality of life in patients with ADPKD.<sup>8</sup> Recent studies have shown that somatostatin analogs can decrease liver volume in patients with ADPKD as well as isolated polycystic liver disease (PLD), and more drug interventions are currently being studied.<sup>9</sup> Here, liver volume is often used as a primary or secondary outcome measure to assess treatment effects on disease progression.

Methods for volume assessment in PKD have improved over time. Historically, estimation methods were used to assess kidney size. Radiologic optimization of volume assessment was introduced by stereology and planimetry of kidney volumes. The Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease study used stereology.<sup>10</sup> To be able to measure small growth differences with low variability, manual planimetry was used in clinical trials investigating treatment effect on kidney growth.<sup>11</sup> Using manual tracing, TKV and total liver volume (TLV) are calculated from a set of contiguous images by summing the products of the slice thickness and the slice area measurements within the kidney or liver boundaries.<sup>10</sup> However, this method is laborious, which limits its use in trial settings and especially, clinical care. With emerging clinical trials that might ameliorate kidney and liver growth in PKD and polycystic livers and even more after introduction of tolvaptan for clinical care, there is an unmet need to have more accurate assessment of kidney as well as liver volumes that can be used quickly and reliably.

Recently, our group developed a deep learning–based approach for fully automated segmentation of polycystic kidneys that performs at a level comparable with interobserver variability.<sup>12</sup> It thus could be considered a replacement for segmentation of polycystic kidneys by a human. Deep learning methods have a unique approach to solving classic image processing tasks by allowing the computer to identify image features of interest. This is in contrast to traditional machine learning that requires predefining the features of interest (e.g., image edges, intensity, and/or texture).

Here, we validate the automated segmentation of polycystic kidneys in an external cohort and add fully automated segmentation of polycystic livers to the deep neural network. The aims of our study are to validate the performance of our fully automated segmentation in an independent external cohort and train our method to additionally generate liver segmentations in these patients with PKD.

## METHODS

### Study Design

This study is a collaboration between the Developing Intervention Strategies to Halt Progression of Autosomal Dominant Polycystic Kidney Disease (DIPAK) Consortium

### Significance Statement

Because renal function often remains stable in the early stages of autosomal dominant polycystic kidney disease (ADPKD), total kidney volume (TKV) is the most important biomarker of disease severity and progression. The authors developed and validated a fully automated method of simultaneously measuring TKV and total liver volume (TLV) as well as their growth rates from magnetic resonance images. Their findings indicate that the method's performance is equivalent to the gold standard of manual tracing while requiring a fraction of the time and skill. This approach may address the unmet need for rapid and reliable assessment of TKV and TLV, allowing identification of patients with rapidly progressing ADPKD for emerging drug treatments and serving as a reliable alternative for assessment of disease progression in clinical trials.

and the Human Imaging Core of the Mayo Clinic PKD Center (Rochester, MN). The Human Imaging Core of the Mayo Clinic PKD Center previously developed a fully automated method of kidney volumetry.<sup>12</sup> This method was further developed to segment individual kidneys and livers, and it is here validated in an external cohort of magnetic resonance imaging (MRI) scans obtained during the DIPAK-1 study, a randomized, controlled, multicenter clinical trial of patients with ADPKD in The Netherlands.<sup>13</sup> ADPKD was diagnosed on the basis of the modified Ravine criteria. The institutional review board at each site approved the protocol. The study was conducted in accordance with the International Conference of Harmonization Good Clinical Practice Guidelines and in adherence to the ethics principles that have their origin in the Declaration of Helsinki. All patients gave written informed consent.

### MRI Data

All participants underwent standardized abdominal MRI scans as part of the DIPAK-1 study.<sup>13</sup> Multiple sequences were scanned, but only the coronal fat saturated T2 single-shot fast spin echo sequence was used for this study. A more detailed description of the MRI protocol has been published previously.<sup>14</sup> The manually traced kidney and liver volumes were assessed as secondary end points for the DIPAK-1 study.<sup>15</sup> The MRIs were made at baseline and after 120 weeks (end of treatment) or 132 weeks (end of study) for longitudinal analyses. DICOM image data from the DIPAK-1 study were transferred to Mayo Clinic after anonymization. More details regarding the MRI dataset can be found in Supplemental Material.

### Reference Standard TKV and TLV

The kidney and liver boundaries were manually traced using Analyze 11.0 (Analyze Direct, Inc., Overland Park, KS). The kidney and liver volumes were calculated from the set of contiguous images by summing the products of the area measurements within the kidney or liver boundaries and slice thickness. Nonrenal parenchyma (e.g., the renal hilum) was excluded from kidney measurement, and the portal vein and gall bladder were excluded from liver measurement. Importantly, all

measurements were performed by readers blinded to patient identity and previous volume measurements.

### Deep Learning Model

We extended our previously developed convolutional neural network architecture<sup>12</sup> to perform semantic segmentation of individual kidneys and livers. A more detailed description of the development and technicalities can be found in Supplemental Material. For training and validation, *K*-fold cross-validation was performed (*K*=5). A total of 440 MRIs were used, and the network was trained on 80% of the images and validated on the remaining 20% for each fold. With this training and validation, the deep learning program learned to measure TLV in addition to TKV, for which it was trained and validated previously.<sup>12</sup> After training, an additional test set of images was used. This included 100 patients, 45 of whom had two MRIs at different time points, which allowed for longitudinal analysis (*i.e.*, a total of 145 images). A flowchart of all MRIs used can be found in Supplemental Figure 1.

### Evaluation of Automated Approach

Comparison statistics were generated from the reference standard segmentations and those made by the automated approach. These comparison statistics are described in detail in Supplemental Material.

### Statistical Analyses

We compared the performance of manually and automated TKV and TLV measurements. First by comparing absolute with percentage differences between volumes measured for right kidney, left kidney, and total kidney as well as TLV. Second, intraclass correlation coefficient (ICC) as well as Bland–Altman analyses were performed to compare the automated measurement method with the reference gold standard manual tracing with bias and precision indicating the observed mean difference and variance (*i.e.*, SD) and 95% limits of agreements. Bland–Altman analyses were performed for cross-sectionally as well as longitudinally analyses. To assess inter-reader variability, we measured 24 MRIs two times. These repeated measurements were performed at random by one of our eight readers. Bland–Altman analyses were also performed for inter-reader variability to compare the variability of manual tracing with the variability between both methods. Lastly, we compared the Mayo Risk classification for all patients while using manually traced TKV with usage of automatically traced TKV. As sensitivity analysis, the performance of TLV for livers >2500 ml was analyzed.

## RESULTS

### Patient Characteristics

The demographics of the patients used for the cross-sectional as well as longitudinal analyses are shown in Table 1. The mean

age of the 100 patients included for the cross-sectional analysis was  $49.1 \pm 7.4$  years old, and 45% were men. Their mean ( $\pm$ SD) eGFR (Chronic Kidney Disease Epidemiology Collaboration equation<sup>16</sup>) was  $49.1 \pm 10.1$  ml/min per  $1.73 \text{ m}^2$ , with a TKV of  $2390 \pm 1585$  ml and a TLV of  $2454 \pm 1188$  ml. The subset of 45 patients used for the longitudinal analysis had similar patient characteristics.

### Optimization of the Artificial Multiobserver Network

Supplemental Figure 2 graphically shows the deep learning network architecture. Similarity metrics are shown in Table 2. The similarity metrics are not significantly different for left versus right kidney (*i.e.*, the algorithm performed equally well for both kidneys). Examples comparing manual with fully automated traced kidneys and livers are given in Supplemental Figure 3A, where we show the largest percentage volume difference in liver when comparing the measurement obtained by manual tracing versus automation, and Supplemental Figure 3B, where we highlight the case with the lowest Dice value.

### Performance of TKV and TLV Assessment

Correlation plots and Bland–Altman analyses comparing the gold standard with our fully automated deep learning approach for TKV as well as TLV are shown in Figure 1, and they are compared with the inter-reader variability of the reference method (manual tracing) of TKV and TLV as shown in Figure 2. For TKV assessment, the bias (mean) and precision (SD) were  $<0.1\%$  (0.006%) and 2.7%, respectively (Figure 1, upper panel). TLV had a slightly larger bias of  $-1.6\%$  but a similar precision of 3.1% (Figure 1, lower panel). Inter-reader variabilities were very similar, with a bias and precision of  $<0.1\%$  (0.03%) and 3.5%, respectively, for TKV and  $-1.5\%$  and 4.8%, respectively, for TLV. Average volumes and percentage differences of the manual versus automated assessment of TKV and TLV are shown in Table 3. Measurements for both right and left kidneys as well as TKV were not significantly different between the reference standard and the automated approach. Liver volumes were significantly different between manual and automated measured volumes, with automated volumes being smaller; absolute difference was  $-31.8 \pm 71.5$  ml ( $P < 0.001$ ), and percentage difference was  $-1.6\% \pm 3.1\%$  ( $P < 0.001$ ). TKV has an accuracy for percentage of measurements within 5% of each other (P5) of 89% and an accuracy of percentage of measurements within 10% of each other (P10) of 100%. TLV has an accuracy of 89% for P5 and an accuracy of 99% for P10. The accuracy of the inter-reader variability for TKV is 86% for P5 (24 of 28) and 100% for P10, whereas it is 75% (six of 24) for P5 and 92% (two of 24) for P10 of TLV. When calculating the Mayo risk classification,<sup>17</sup> only two of 100 patients were reclassified into another risk class (Table 4). The details of the two reclassified subjects are provided in Supplemental Table 1 and show that one was almost classified in the same risk category (patient 1), where the automated method gave

**Table 1.** Patient characteristics

Characteristic	Cross-Sectional, n=100	Longitudinal, n=45
Age, yr	49.1±7.4	49.1±6.7
Men	45 (45)	20 (44.4)
Body mass index, kg/m <sup>2</sup>	26.5±4.1	26.4±3.5
Body surface area, m <sup>2</sup>	1.98±0.20	1.99±0.16
Length, m	1.76±0.10	1.76±0.09
Systolic BP, mm Hg	132.8±14.2	132.8±15.8
Diastolic BP, mm Hg	82.1±9.6	81.6±9.2
Antihypertensive medication	88 (88)	41 (91.1)
Plasma creatinine, μmol/L	130±32	129±31
eGFR, CKD-EPI, ml/min per 1.73 m <sup>2</sup>	49.1±10.1	49.5±10.2
24-h urine volume, L	2.3±0.9	2.3±0.8
Albuminuria, mg/24 h	48.4 [17.3–70.4]	55.0 [18.4–80.2]
Kidney volume, ml		
Total	2390±1585	2194±1311
Left	1223±846	1139±702
Right	1168±771	1055±640
Liver volume, ml	2454±1188	2368±1059

Values for categorical variables are given as number (percentage); values for continuous variables are given as mean ± SD or median (interquartile range). CKD-EPI, Chronic Kidney Disease Epidemiology Collaboration.

a slightly too low height-adjusted TKV (283.33, where 284.53 was the cutoff point for the same risk class as the manual method).

When only analyzing livers with a volume of >2500 ml, the automated TLV was 3802±1397, and manual TLV was 3826±1404 ml; these values were not significantly different when analyzed in absolute as well as percentage difference ( $P=0.20$  and  $P=0.10$ , respectively). The accuracies for this subgroup were 93% for P5 and 100% for P10.

### Performance of the Fully Automated Method in Detecting Kidney and Liver Growth

Table 5 shows the absolute as well as percentage observed growth of each kidney, TKV, and TLV. No significant differences between the manually traced and fully automated growth rates were observed for any of the kidney volumes. However, liver growth in absolute as well as percentage growth differed significantly ( $P<0.01$  and  $P=0.002$ , respectively). Figure 3 shows Bland–Altman analyses of our fully automated deep learning approach in detecting the growth of kidneys and liver in PKD. Bias (mean) and precision (SD) of TKV

assessment were <0.1% (0.05%) and 3.2%, respectively. TLV had bias and mean of 1.4% and 2.9%, respectively.

When only analyzing the subgroup of enlarged livers (>2500 ml), no significant difference was observed with absolute growth of 53±396 ml for manual TLV and 75±385 ml for automated TLV ( $P=0.30$ ) or percentage growth rate, being 2.7%±11.8% and 3.4%±11.4%, respectively ( $P=0.30$ ).

## DISCUSSION

This study introduces our fully automated segmentation artificial deep neural network for a combined kidney and liver volume assessment and validates this method in an external cohort with polycystic kidneys and livers. It strengthens the observation that this fully automated method, without

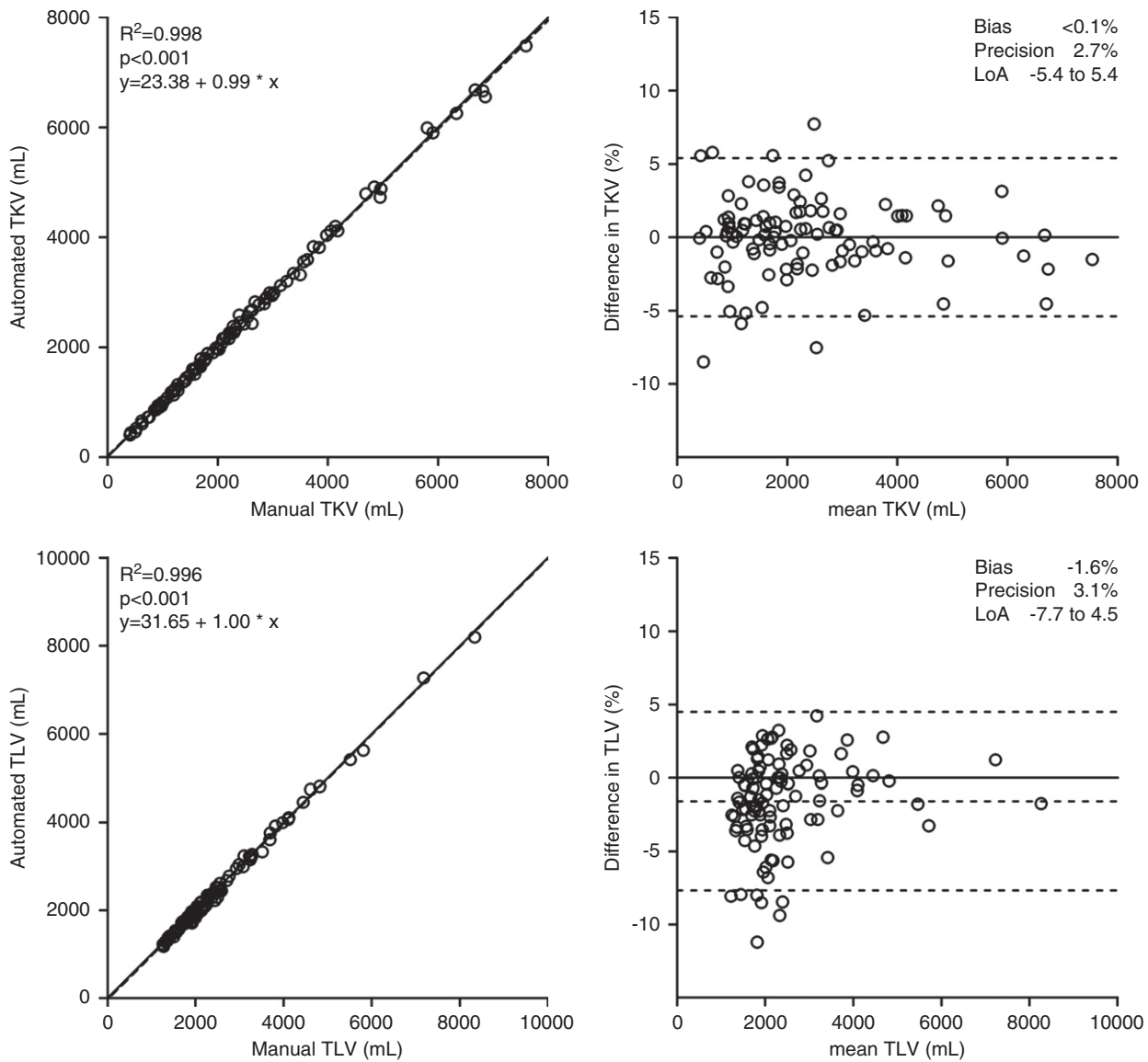
the interference of any human action, performs at a level comparable with the variability observed when two different people measure TKV or TLV in ADPKD. Automated measurement has the ability to detect similar kidney growth rates compared with manual tracing and only leads to 2% reclassification of the Mayo risk classification. Furthermore, this study introduces a fully automated total liver segmentation that performs as accurately as manually traced liver volumes measured by two people.

TKV has been shown to predict CKD progression in ADPKD,<sup>1</sup> and it has been shown to increase in early disease; however, GFR remains stable for a long time while the disease is already progressing.<sup>1,3</sup> Also, higher rates of TKV increase have been shown to be associated with more rapid GFR decline,<sup>3</sup> and higher baseline TKV has been shown to be associated with more rapid TKV growth.<sup>3,17</sup> Height-adjusted TKV at a given age is clinically used to assess the rapidness of disease progression: the Mayo Clinic risk classification for ADPKD developed by Irazabal *et al.*<sup>17</sup> Kidney volume and growth are recognized in PKD as a marker for disease severity and progression,<sup>5,6</sup> and they are—together

**Table 2.** Summary statistics for the automated approach compared with the gold standard

Similarity Metric	Left Kidney	Right Kidney	Total Kidney Volume	Liver Volume
Jaccard	0.92±0.03 [0.82/0.97]	0.91±0.04 [0.79/0.97]	0.92±0.03 [0.80/0.97]	0.90±0.03 [0.75/0.97]
Dice	0.96±0.02 [0.90/0.99]	0.95±0.02 [0.88/0.98]	0.96±0.02 [0.89/0.98]	0.95±0.02 [0.86/0.98]
Sensitivity	0.96±0.03 [0.88/0.98]	0.99±0.02 [0.84/0.99]	0.96±0.02 [0.87/0.99]	0.94±0.03 [0.85/0.98]
Precision	0.96±0.02 [0.90/0.99]	0.99±0.03 [0.84/0.99]	0.96±0.02 [0.87/0.99]	0.96±0.02 [0.87/0.99]
Mean surface distance, mm	0.48±0.27 [0.16/1.72]	0.56±0.33 [0.16/1.87]	0.52±0.30 [0.16/1.79]	0.71±0.38 [0.17/2.92]
Hausdorff distance, mm	7.04±4.70 [1.90/23.1]	7.89±5.52 [2.13/25.6]	8.71±6.04 [2.30/25.4]	12.0±6.34 [2.82/27.1]

Shown are the results for total kidney volumes as well as the liver volumes (data are given as mean ± SD, with minimum/maximum values between brackets). Mean surface distance and Hausdorff distance are reported in terms of voxel distance.



**Figure 1.** High correlation (left panels) and agreement, assessed using Bland–Altman (right panels) for total kidney volume (TKV) measurements (upper panels) and total liver volume (TLV) measurements (lower panels) obtained by the automated approach and manual segmentation. In the left panels, the solid lines represent no difference, whereas the dotted lines show the actual regression line; in the right panels, the solid lines show the actual mean difference (bias), and the dotted lines show 95% limits of agreements (LoAs).

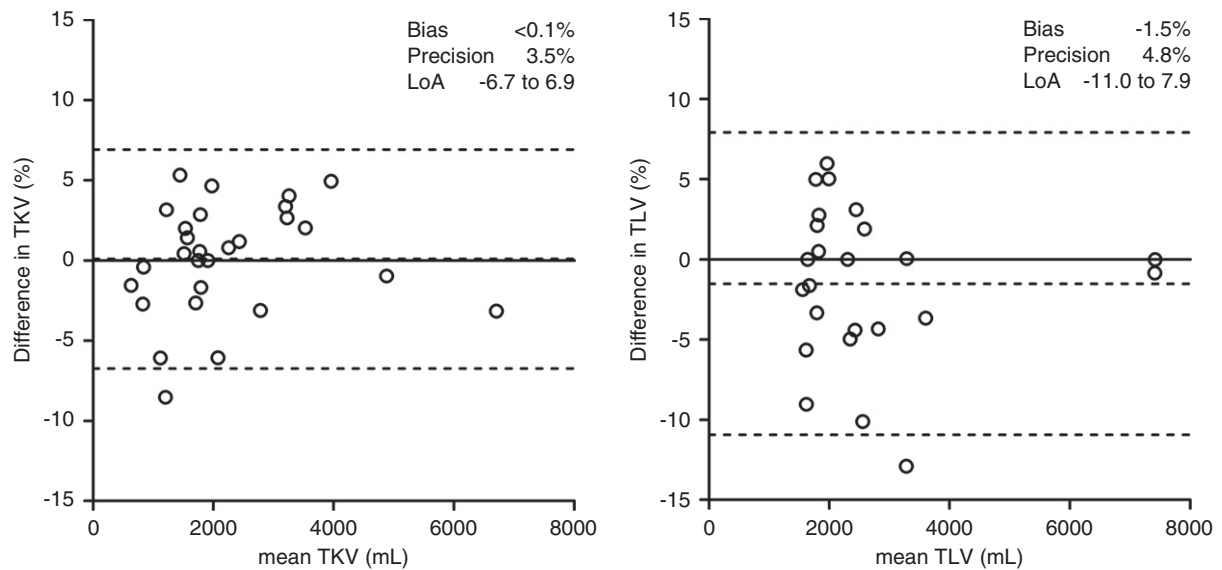
with GFR—often used as primary or secondary outcome measures in research trials.

For automated kidney segmentations in PKD, here we validated the previously published automated method developed by our research group in an independent dataset of patients with ADPKD in The Netherlands that independently measured TKVs and TLVs at different centers with different image acquisition protocols.<sup>12</sup> This strengthens its potential as replacement for manual tracing as assessment of TKV. Comparing the volumes observed by the deep learning method with inter-reader variability of manual tracing, the gold standard, no differences in accuracy are found. Also, only two of 100 patients were reclassified to

another Mayo risk class, which is clinically used to select patients with rapid disease progression who will benefit the most from drug treatments and thus, are selected for treatment or clinical trials.

When analyzing change in TKV, automatically measured kidney volumes show the same rate of change, making them a potential replacement for manual tracing to follow disease progression. Our automated measurement of TKV is the first method to have no difference in rate of change compared with manual tracing, and therefore, performs even better compared with manual measurement of growth using two different MRI sequences, the T1 three-dimensional spoiled gradient echo and T2 single-shot fast spin echo, that were





**Figure 2.** The agreement, using Bland–Altman, of the reference method (manual tracing) between two different readers is comparable to the agreement between manual tracing and automated segmentation (Figure 1), for both total kidney (left panel) as well as liver volume (right panel). In the left panel, the solid line represents no difference, whereas the dotted lines show the actual regression line; in the right panel, the solid line represents the actual mean difference (bias), and the dotted lines show 95% limits of agreement (LoAs).

previously shown to have significantly different absolute growth rates.<sup>18</sup> Currently, there is no recommendation regarding usage of only T1-weighted or only T2-weighted images to assess kidney growth in ADPKD. However, a recent study detailed recommendations for standardization of TKV measurements in longitudinal studies.<sup>19</sup> Another study showed that the quality of T1 images is more often of too low quality to use for assessment of kidney and liver volumes.<sup>20</sup> Furthermore, it is more difficult to identify the borders of the kidneys on T1-weighted images, leading to increased inter-reader variability, which will introduce an increased variability in the deep learning program, and we were afraid that we would introduce a higher variability. Future studies could test the performance of the deep learning program on T1-weighted images. Automated measurement eliminates the element of inter-reader variability. Measurement of change in TKV using the automated method, therefore, holds promise as replacement of manual tracing of

TKVs for growth analyses. This could be used in clinical care to assess the individual growth rate of a patient, but moreover, it could be used in clinical trials to assess kidney growth as an outcome parameter.

Fortunately, not all patients with PKD have polycystic livers. However, studies have previously reported the disease prevalence ranging from 69.3% and 78.7% in men and women, respectively, with early-stage ADPKD<sup>8</sup> to a prevalence of over 94% for lifelong development of at least one liver cyst.<sup>7</sup> The severity of the polycystic livers has been shown to be negatively correlated with quality of life.<sup>8</sup> In recent years, somatostatin analogs have been shown to decrease liver volumes in patients with PLD as well as patients with ADPKD who have polycystic livers, and more therapies are being developed.<sup>9</sup> Liver volume and its growth are important markers to evaluate treatment efficacy, and they are often used as a primary or secondary end points in clinical trials.

**Table 3.** Differences in kidney and liver volumes when measured using manually traced or automated segmentation

Organ	Differences in Volume (Manual Versus Automated Measurement)									
	Volumes (ml)		Absolute (ml)			Percentage (%)			Accuracy (%)	
	Automated	Manual	Bias	Precision	P Value	Bias	Precision	P Value	P5	P10
Left kidney	1220±839	1223±846	-2.9	42.5	0.80	-0.2	3.0	0.50	91	99
Right kidney	1168±761	1168±771	0.1	40.4	0.30	0.2	3.5	0.50	85	99
Total kidney	2388±1569	2390±1585	-2.8	70.7	0.80	0.0	2.8	0.90	89	100
Liver	2422±1190	2454±1188	-31.8	71.5	<0.001	-1.6	3.1	<0.001	86	99

Values are given as mean ± SD. P values are calculated using a paired Wilcoxon signed rank test for absolute differences; for percentage differences, a one-sample t test was used. P5 and P10 represent the percentages of measurements within 5% and 10% of each other, respectively.

**Table 4.** Automated versus manual assessed total kidney volume on stratification according to Mayo height-adjusted TKV risk classes (A, lowest risk; E, highest risk)

Automated TKV	Manual TKV				
	A	B	C	D	E
A	3	0	0	0	0
B	1	17	0	0	0
C	0	0	32	1	0
D	0	0	0	29	0
E	0	0	0	0	12

Information regarding the two misclassified cases is also given. TKV, total kidney volume.

Liver segmentation and assessment of volume are used in polycystic livers that are part of ADPKD, but they are also used in PLD and for various other indications. For instance, it is measured before major hepatectomy, portal vein embolization, and liver transplantation. The gold standard for liver volume assessment is, like for TKV assessment, manually tracing the liver. Recently, both semiautomated and fully automated methods have been developed for segmentation of livers.<sup>18</sup> As far as we know, only one study performed automated segmentations of polycystic livers, where a correlation (assessed as intraclass correlation coefficient) of 0.91 for manually traced compared with automatically segmented liver volumes in livers ranging from 1 to 5 L in volume was achieved.<sup>21</sup> This previously reported ICC is lower than our finding (ICC of 0.996) for automated liver volume assessment. In our cohort, we observed a significant difference in liver volume and liver growth in absolute (milliliters) as well as percentage growth. However, significantly different, these differences were numerically small, with means ( $\pm$ SD) of  $2422 \pm 1190$  and  $2454 \pm 1188$  for automated and manual tracing, respectively: a difference of 1.3%. Moreover, the precision and bias when comparing automated with manual tracing were comparable with those observed when two individual readers traced the liver volumes manually (inter-reader variability), with precision and bias of  $-1.6\% \pm 3.1\%$  and  $-1.5\% \pm 4.8\%$ , respectively. Furthermore, we observed that liver volumes were more reliably measured when patients suffered from moderate to severe PLD (liver volumes  $>2500$  ml). This holds promise for the use of our automated liver volume measurement, because it is as accurate as manual tracing, and the method seems to be more accurate for the livers for which it is most relevant to require measurement (*i.e.*, those with cysts). A future study focusing on polycystic liver volume assessment specifically could optimize our method.

We believe that an automated approach of polycystic kidney and liver segmentation that performs equally to manually traced kidney and liver volumes has advantages over manually traced kidneys. It reduces the costs of the laborious measurement. It has the advantage of being a quick and easily applicable method that can be used *ad hoc*—results are computed in a matter of seconds versus 60–120 minutes for manually traced kidney and liver segmentations—and without the need to have

**Table 5.** Change in kidney and liver volume during 2.5 years of follow-up when measured using manually traced or automated segmentation

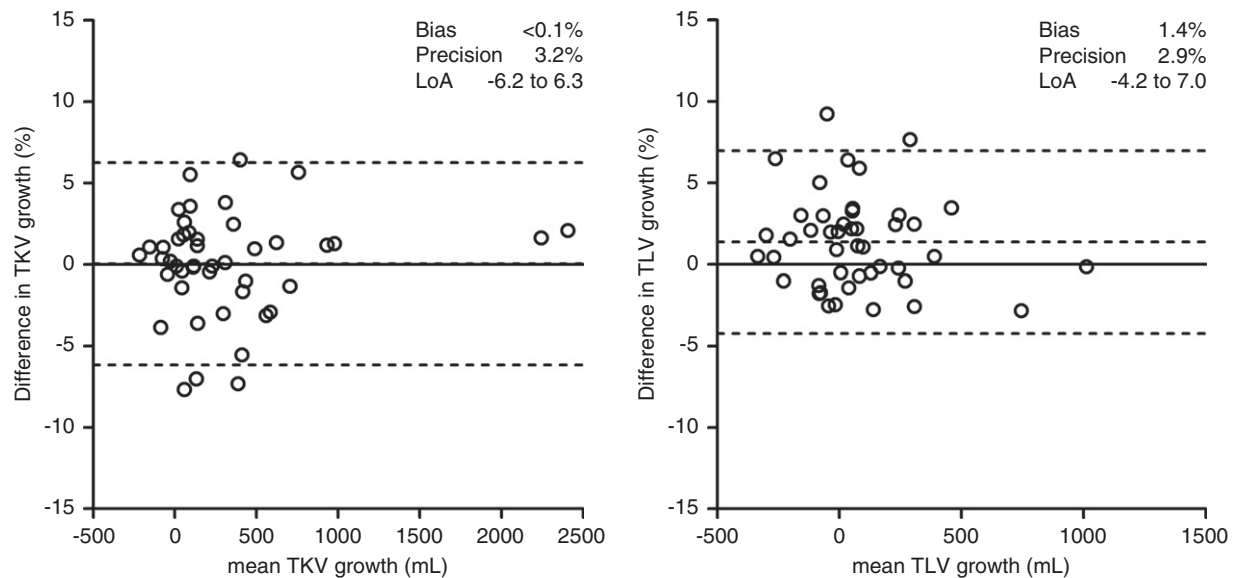
Organ	Manual	Automated	P Value
Left kidney			
Change, ml	132 $\pm$ 195	138 $\pm$ 202	0.20
Change, %	10.6 $\pm$ 12.9	11.1 $\pm$ 13.3	0.30
Right kidney			
Change, ml	194 $\pm$ 389	191 $\pm$ 377	0.60
Change, %	13.4 $\pm$ 18.8	13.1 $\pm$ 18.4	0.50
Total kidney			
Change, ml	325 $\pm$ 514	329 $\pm$ 522	0.70
Change, %	12.1 $\pm$ 15.0	12.2 $\pm$ 15.0	0.90
Liver			
Change, ml	56 $\pm$ 262	85 $\pm$ 251	<0.01
Change, %	2.7 $\pm$ 10.4	4.1 $\pm$ 9.9	0.002

Values are given as mean  $\pm$  SD. P values were calculated using a paired t test.

any expertise in segmentation, and thereby, it provides access to a much broader group of physicians and scientists. Because it is as accurate as multiobserver human kidney volume measurements, it can be applied both in research trials as well as in clinical care, and it has the advantage that it will recognize the same tissue as kidney versus surrounding tissue on multiple occasions of measurement. This reduces inpatient variability, especially in assessing growth of kidneys and liver over time, which was also indicated by the fact that the automated method can measure kidney growth as adequately as manually traced segmentations. It is our future goal to make our method publicly available.

The strengths of our study are that we validated and optimized the performance of our previously developed fully automated segmentation artificial deep neural network in a separate cohort of patients with ADPKD whose TKVs and TLVs were assessed by manually tracing in a different institute and scanned under a different protocol (*i.e.*, different magnetic resonance pulse sequence), which demonstrates its robustness. Furthermore, this is the first study to combine total kidney with liver volume assessment in patients with ADPKD and the first to measure change in TKV and TLV over time using automated volume assessment. This program can also be used for PKD with other etiologies as well as PLD. A future study including only patients with PLD would enable us to optimize and validate this automated method for liver volume assessment specifically in PLD. A potential weakness of our study is the relatively low number of patients with moderate to severe polycystic liver enlargement. Also, the average quality of the MRIs obtained during the DIPAK-1 study might be higher than the average quality of MRIs in clinical care, because kidney and liver volumes were secondary end points in this study.

In conclusion, we validated our fully automated segmentation approach for TKV measurement in ADPKD. In this study, we additionally trained and validated our approach for TLV measurement. This method performs equally to the



**Figure 3.** Agreement between growth rates of total kidney (left panel) as well as liver volume (right panel) is very high for the automated approach compared to manual segmentation. In the left panel, the solid line represents no difference, whereas the dotted lines show the actual regression line; in the right panel, the solid line represents the actual mean difference (bias), and the dotted lines show 95% limits of agreement (LoAs).

gold standard of manual tracing for both TKV and TLV measurements. This provides a useful tool for the assessment of disease severity as well as treatment effects in clinical care, where now inferior estimations are available. It furthermore holds promise as an alternative to laborious, dedicated, and expensive manual tracing that currently is performed in clinical trials to assess kidney volume and growth. With this validation, we conclude that our method provides a reliable alternative to manually tracing TKV and TLV in patients with ADPKD.

## ACKNOWLEDGMENTS

The Developing Intervention Strategies to Halt Progression of Autosomal Dominant Polycystic Kidney Disease (DIPAK) Consortium is an interuniversity collaboration in The Netherlands that was established to study autosomal dominant polycystic kidney disease and develop rational treatment strategies for this disease. For this study, no additional funding was received.

## DISCLOSURES

None.

## FUNDING

This work was supported in part by the Mayo Clinic Robert M. and Billie Kelley Pirnie Translational PKD Center and National Institute of Diabetes and Digestive and Kidney Diseases grants P30DK090728 and K01DK110136 as well as PKD Foundation grant 206g16a. The DIPAK Consortium is sponsored by Dutch

Kidney Foundation grants CP10.12 and CP15.01 and Dutch Government grant LSHM15018.

## SUPPLEMENTAL MATERIAL

This article contains the following supplemental material online at <http://jasn.asnjournals.org/lookup/suppl/doi:10.1681/ASN.2018090902/-/DCSupplemental>.

Supplemental Figure 1. Flowchart of all MRIs used.

Supplemental Figure 2. Schematic of the deep neural network architecture developed in this study.

Supplemental Figure 3. Examples of automated segmentations highlighting two cases: the case with the largest volume difference for liver within one patient (between manual and automated) and the case with the lowest liver Dice score, which is a measure of accuracy and overlap between both methods, calculated using the amount of voxels that were positive for both methods (true positives) and the amount of voxels that were false negatives (the automated method did not identify a voxel, whereas the gold standard method did) and false positives (vice versa).

Supplemental Material. Methods.

Supplemental Table 1. Automated versus manual assessed TKV on stratification according to Mayo height-adjusted TKV risk classes (A, lowest risk; E, highest risk).

## REFERENCES

- Chapman AB, Bost JE, Torres VE, Guay-Woodford L, Bae KT, Landsittel D, et al.: Kidney volume and functional outcomes in autosomal dominant polycystic kidney disease. *Clin J Am Soc Nephrol* 7: 479–486, 2012



2. Higashihara E, Nutahara K, Okegawa T, Shishido T, Tanbo M, Kobayasi K, et al.: Kidney volume and function in autosomal dominant polycystic kidney disease. *Clin Exp Nephrol* 18: 157–165, 2014
3. Grantham JJ, Torres VE, Chapman AB, Guay-Woodford LM, Bae KT, King BF Jr, et al.: CRISP Investigators: Volume progression in polycystic kidney disease. *N Engl J Med* 354: 2122–2130, 2006
4. Fick-Brosnahan GM, Belz MM, McFann KK, Johnson AM, Schrier RW: Relationship between renal volume growth and renal function in autosomal dominant polycystic kidney disease: A longitudinal study. *Am J Kidney Dis* 39: 1127–1134, 2002
5. FDA: Qualification of biomarker, total kidney volume in studies for treatment of autosomal dominant polycystic kidney disease, 2018. Available at: <https://www.fda.gov/downloads/Drugs/Guidances/UCM458483.pdf12/13/2018>. Accessed July 20, 2018
6. EMA: Qualification opinion, Total Kidney Volume (TKV) as a prognostic biomarker for use in clinical trials evaluating patients with Autosomal Dominant Polycystic Kidney Disease (ADPKD), 2018. Available at: [http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Regulatory\\_and\\_procedural\\_guideline/2015/11/WC500196569.pdf12/13/2018](http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2015/11/WC500196569.pdf12/13/2018). Accessed July 20, 2018
7. Bae KT, Zhu F, Chapman AB, Torres VE, Grantham JJ, Guay-Woodford LM, et al.: Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease (CRISP): Magnetic resonance imaging evaluation of hepatic cysts in early autosomal-dominant polycystic kidney disease: The consortium for radiologic imaging studies of polycystic kidney disease cohort. *Clin J Am Soc Nephrol* 1: 64–69, 2006
8. Hogan MC, Abebe K, Torres VE, Chapman AB, Bae KT, Tao C, et al.: Liver involvement in early autosomal-dominant polycystic kidney disease. *Clin Gastroenterol Hepatol* 13: 155–64.e6, 2015
9. van Aerts RMM, van de Laarschot LFM, Banales JM, Drenth JPH: Clinical management of polycystic liver disease. *J Hepatol* 68: 827–837, 2018
10. Chapman AB, Guay-Woodford LM, Grantham JJ, Torres VE, Bae KT, Baumgarten DA, et al.: Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease cohort: Renal structure in early autosomal-dominant polycystic kidney disease (ADPKD): The Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease (CRISP) cohort. *Kidney Int* 64: 1035–1045, 2003
11. King BF, Reed JE, Bergstralh EJ, Sheedy PF 2nd, Torres VE: Quantification and longitudinal trends of kidney, renal cyst, and renal parenchyma volumes in autosomal dominant polycystic kidney disease. *J Am Soc Nephrol* 11: 1505–1511, 2000
12. Kline TL, Korfiatis P, Edwards ME, Blais JD, Czerwiec FS, Harris PC, et al.: Performance of an artificial multi-observer deep neural network for fully automated segmentation of polycystic kidneys. *J Digit Imaging* 30: 442–448, 2017
13. Meijer E, Drenth JP, d'Agnolo H, Casteleijn NF, de Fijter JW, Gevers TJ, et al.: DIPAK Consortium: Rationale and design of the DIPAK 1 study: A randomized controlled clinical trial assessing the efficacy of lanreotide to Halt disease progression in autosomal dominant polycystic kidney disease. *Am J Kidney Dis* 63: 446–455, 2014
14. Spithoven EM, van Gastel MD, Messchendorp AL, Casteleijn NF, Drenth JP, Gaillard CA, et al.: DIPAK Consortium; DIPAK Consortium: Estimation of total kidney volume in autosomal dominant polycystic kidney disease. *Am J Kidney Dis* 66: 792–801, 2015
15. Meijer E, Visser FW, van Aerts RMM, Blijdorp CJ, Casteleijn NF, D'Agnolo HMA, et al.: DIPAK-1 Investigators: Effect of lanreotide on kidney function in patients with autosomal dominant polycystic kidney disease: The DIPAK 1 randomized clinical trial. *JAMA* 320: 2010–2019, 2018
16. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF 3rd, Feldman HI, et al.: CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration): A new equation to estimate glomerular filtration rate. *Ann Intern Med* 150: 604–612, 2009
17. Irazabal MV, Rangel LJ, Bergstralh EJ, Osborn SL, Harmon AJ, Sundsbak JL, et al.: CRISP Investigators: Imaging classification of autosomal dominant polycystic kidney disease: A simple model for selecting patients for clinical trials. *J Am Soc Nephrol* 26: 160–172, 2015
18. Gotra A, Sivakumaran L, Chartrand G, Vu KN, Vandembroucke-Menu F, Kauffmann C, et al.: Liver segmentation: Indications, techniques and future directions. *Insights Imaging* 8: 377–392, 2017
19. Edwards ME, Blais JD, Czerwiec FS, Erickson BJ, Torres VE, Kline TL: Standardizing total kidney volume measurements for clinical trials of autosomal dominant polycystic kidney disease. *Clin Kidney J* 12: 71–77, 2018
20. van Gastel MDA, Messchendorp AL, Kappert P, Kaatee MA, de Jong M, Renken RJ, et al.: DIPAK Consortium: T1 vs. T2 weighted magnetic resonance imaging to assess total kidney volume in patients with autosomal dominant polycystic kidney disease. *Abdom Radiol (NY)* 43: 1215–1222, 2018
21. Kim Y, Bae SK, Cheng T, Tao C, Ge Y, Chapman AB, et al.: Automated segmentation of liver and liver cysts from bounded abdominal MR images in patients with autosomal dominant polycystic kidney disease. *Phys Med Biol* 61: 7864–7880, 2016