
Article:
Simms, R.J., Doshi, T., Metherall, P. et al. (7 more authors) (2019) A rapid high-performance semi-automated tool to measure total kidney volume from MRI in autosomal dominant polycystic kidney disease. European Radiology. ISSN 0938-7994

https://doi.org/10.1007/s00330-018-5918-9

A rapid high performance semi-automated tool to measure Total Kidney Volume from MRI in Autosomal Dominant Polycystic Kidney Disease

Roslyn J Simms PhD[1,2,3*], Trushali Doshi PhD[1*], Peter Metherall PhD[3,4], Desmond Ryan PhD[1], Peter Wright PhD[4], Nicolas Gruel PhD[3], Maatje D.A. van Gastel PhD[5], Ron T Gansevoort PhD[5], Wendy Tindale PhD[3,4], Albert CM Ong DM[1,2,3]

[1]Kidney Genetics Group, Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield, UK.

[2]Sheffield Kidney Institute, Sheffield Teaching Hospitals NHS Foundation Trust, UK.

[3]Institute for in silico medicine, University of Sheffield, UK.

[4]Medical Imaging and Medical Physics, Sheffield Teaching Hospitals NHS Foundation Trust, UK.

[5]Department of Nephrology, University Medical Centre Groningen, the Netherlands

*equal contribution

Corresponding author

Prof Albert CM Ong

Kidney Genetics Group, Academic Unit of Nephrology, Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield Medical School, Beech Hill Road, Sheffield S10 2RX, UK.

Tel.: 0114-215-9542

Fax: 0114-271-3892

Email: a.ong@sheffield.ac.uk

Abstract

Objectives: To develop a new semi-automated method (Sheffield TKV Tool) for measuring total kidney volume (TKV) from magnetic resonance images (MRI) in patients with autosomal dominant polycystic kidney (ADPKD).

Methods: TKV was initially measured in 61 patients with ADPKD using the Sheffield TKV Tool and its performance compared to manual segmentation and other published methods (ellipsoidal, mid-slice, MIROS). It was then validated using an external dataset of 65 MRI scans.

Results: 61 patients (mean age 45±14years, baseline eGFR 76±32ml/min/1.73m^2) with ADPKD had a wide range of TKV (258-3680ml) measured manually. The Sheffield TKV Tool was highly accurate (mean volume error of 0.5±5.3% for right kidney, -0.7±5.5% for left kidney), reproducible (intra-operator variability -0.2±1.3% and inter-operator variability of 1.1±2.9%) and outperformed published methods. It took less than 6 minutes to execute and performed consistently in an external MRI dataset with TKV acquired using three different scanners and measured using a different segmentation methodology (mean volume error was 3.45±3.96%, n=65).

Conclusions: The Sheffield TKV Tool is operator friendly and requires minimal user interaction. Its rapidity, accuracy and reproducibility highlights it's potential for adoption into general use for measuring TKV in ADPKD.

Keywords (max 5 keywords)

Polycystic kidney diseases; ADPKD; Magnetic resonance imaging; kidneys

Key points (3 to 5 key points)

This new semi-automated method (Sheffield TKV Tool) to measure total kidney volume (TKV) will facilitate the routine clinical assessment of patients with ADPKD.

Abbreviations and acronyms (Non-commonly used)

CoV:	Coefficient of Variation
DSC:	Dice Similarity Coefficient
HtTKV:	Height Adjusted Total Kidney Volume
LKV:	Left Kidney Volume
LSM:	Level Set Method
MIROS:	Minimal Interaction Rapid Organ Segmentation
RKV:	Right Kidney Volume
TRUFI:	True Fast Imaging with Steady-State Free Precession

Introduction

Autosomal dominant polycystic kidney disease (ADPKD) is the commonest inherited kidney disease and fourth leading cause of end stage renal failure (ESRF) worldwide [1, 2]. It is characterised by the gradual progressive development and growth of renal cysts which result in increased total kidney volume (TKV).

Changes in estimated glomerular filtration rate (eGFR) are conventionally used to measure loss of kidney function. In ADPKD however, eGFR does not change until the later stages of disease due to compensatory glomerular hyperfiltration [3] thus limiting its use to late disease [4]. In the earlier stages of disease, increases in TKV are detectable before decreases in eGFR [5]. A single baseline TKV measurement in combination with age and eGFR (Mayo Imaging Classification) has also been shown to accurately predict future decline in kidney function [6]. TKV has been approved by both the US Food and Drug Administration (FDA) and the European Medicines Agency (EMA) as a prognostic biomarker for disease progression in ADPKD to facilitate the enrichment of patients at higher risk of rapid progression in future clinical trials [4]. In addition, Tolvaptan [7] has been approved for use in ADPKD patients in Europe with evidence of rapid disease progression. Guidance from the ERA-EDTA recommends the use of TKV to select higher risk patients for treatment [8].

The current gold standard method for measuring TKV from MRI involves manual tracing of the kidney boundary on each MRI slice using dedicated software and summing the product of area measurements and slice thickness [9]. This is time consuming and subject to intra- and inter-operator variability errors. There is a clear need to develop more rapid and accurate methods for measuring TKV to facilitate its wider adoption into clinical use.

Several semi-automated methods and estimating equations have been developed to address the challenge of measuring TKV in ADPKD [10] (summarised in Supp Table 1). Fully automatic

methods to estimate TKV have also been reported [11-13] although they require a good training dataset to include severely cystic kidneys due to the associated geometric and anatomical variability.

In this paper, we describe the development and validation of a new, rapid, high performance semi-automated method (Sheffield TKV Tool) for measuring TKV in a representative group of patients with ADPKD and a wide range of TKV.

Materials and Methods

Study population

Sixty-one patients with ADPKD and stage 1-3 chronic kidney disease (CKD) attending a specialist PKD clinic at Sheffield Kidney Institute consented to an unenhanced abdominal MRI for measurement of TKV. Renal function (eGFR [14]) was measured at baseline (within one month of the MRI) and the most recent follow up result was recorded (2.00±0.52 years; 0.07-2.72 years). The study was approved by a research ethics committee (13/YH/026).

MRI Acquisition

Kidney MRI scans were coronal true fast imaging with steady-state free precession (TRUFI) T2-weighted sequences (Siemens Avanto 1.5T scanner) with the following parameters: 4mm slice thickness, 0mm slice gap, 2ms echo time, 3.99ms repetition time, 60^{o} flip angle, 0.68x0.68mm in plane resolution and 512x512 acquisition matrix.

Sheffield TKV Tool Development

The Sheffield TKV Tool was implemented using a MATLAB 2016b (MathWorks, Natick, MA) framework. The right or left kidney was segmented individually from coronal MRI slices using image processing techniques (Figure 1). Coronal kidney region slices were selected using mid-sagittal plane (Figure 1: Step 2). Prior to segmentation pre-processing steps were applied to reduce motion artefact, intra- and inter- slice intensity variations [15] from the selected coronal slices (Figure 1: Step 3). On the cropped region of interest (Figure 1: Step 4), the user defined contour (Figure 1: Step 5) was propagated using hybrid level set method (LSM) [16] that utilize edge (gradient) as well as regional statistics to obtain final segmentation boundary. The energy function $\mathrm{E}(\phi)$ used in hybrid LSM [16] is given as:

$$\mathrm{E}(\phi) = -\alpha \int_{\Omega} (\mathrm{I} - \mathrm{u})\mathrm{H}(\phi)\mathrm{d}\Omega + \beta \int_{\Omega} \mathrm{g}\left|\nabla \mathrm{H}(\phi)\right|\mathrm{d}\Omega$$

where ϕ is the level set function. I is the image to be segmented, $\mathrm{H}(\phi)$ is smooth Heaviside function which considers area around contour, g is image edge (boundary) map where contour should be attracted and is set to be $\mathrm{g} = \frac{1}{1 - \mathrm{c}\left|\nabla \mathrm{I}\right|^2}$ with c controlling the slope. Ω is the image domain and α and β are predefined weights to balance two terms. The first term on the right hand side of equation defines that region to be segmented should have intensity greater than μ which is set to 50. Parameters α and β are set to 0.01 and 100 respectively. Evolving contour (level set function) is stopped after 100 iterations to obtain expected kidney region outline (Figure 1: Step 6). These parameters were determined experimentally from a random dataset of 10. The final contour obtained is not highly sensitive to the choice of parameters. After segmentation, kidney volume (KV) was calculated by summing the product of areas of the kidney region and slice thickness. The Sheffield TKV Tool was applied separately for right (R) and left (L) kidneys to enable errors specific to either side to be identified quickly. The tool

was developed and optimised using 10 random cases from the patient cohort (training set) and internally validated on the remaining (51) patients images.

TKV measurements

To obtain reference TKV measurements, the gold standard method of manual segmentation was performed using MIM Maestro (Cleveland, OH, USA) on T2 TRUFI coronal MR images of all 61 patients by an experienced image analyst (PM). PM was blinded to the development and TKV measurements of Sheffield TKV Tool. Consistent with standard methods of manually measuring TKV [9, 15], blood vessels in the kidney and hilum (structures including ureter, blood vessels and nerves entering each kidney) were excluded [17]. A second image analyst (TD) used the Sheffield TKV Tool (TD1) to measure the right (R) and left (L) kidney volumes (KV) separately and compared its performance to that of the mid-slice method [6, 18] and MIROS tool [15] in all 61 patients and used the ellipsoid formula on 51 (typical, class 1) patients. TKV was obtained by summing right and left KV. To apply the MIROS tool, open source code was obtained from https://gitlab.com/Philbrick/rilcontour and re-written in MATLAB. The MIROS tool was developed for HASTE sequences. To enable use on TRUFI sequence, parameters were optimised using 10 random representative cases from the patient cohort (training set) and validated on the remaining (51) patients images.

To assess the inter-operator variability (PM and TD) of manual segmentation, TD manually (TDm1) measured TKV for 40 kidneys from a representative sub-set of 20 patients (TKV 258-3680 ml). TD repeated the manual segmentation on the same dataset after one month (TDm2) to assess the intra-operator variability.

To assess the inter-operator variability of the Sheffield TKV Tool, image analyst PM calculated TKV for a similar dataset of 20 patients (PMs). The intra-operator variability of the Sheffield TKV Tool was assessed by TD on two occasions (TD1, TD2) separated by one month.

Mayo risk classification

We classified patients in the development cohort into Class 1 (typical) or Class 2 (atypical) based on their kidney morphology on MRI as defined in the Mayo Imaging classification ([6]). Class 1 patients were further subdivided into categories 1A-1E, which has been shown to correlate with the rate of disease progression measured by eGFR change ([6]).

Validation of Sheffield TKV Tool

TRUFI MRI renal images of 65 ADPKD patients who participated in the DIPAK-1 study were acquired using one of three scanners (GE Medical Systems (16), Siemens (37) and Philips Healthcare (12)). De-identified DICOM image data from the DIPAK-1 study was transferred to Groningen Medical Centre and converted to the NIFTI file format by the dcm2nii software. The images had a reconstructed matrix size of at least 256 × 256 × Z (with Z large enough to cover the full extent of the kidneys within the imaged volume). Image voxel sizes were most commonly on the order of 1.5 mm in-plane with 4 mm slice thickness and spacing between slices. The Medical Ethics Committee of University Medical Center Groningen approved the trial protocol that 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.

Kidney boundaries were manually traced using commercially available software Analyze Direct 11.0 (Analyze Direct, Inc., Overland Park, KS, USA) and kidney volumes were calculated from the set of contiguous images by summing the products of the area measurements within the kidney boundaries and slice thickness. Non-renal parenchyma, e.g. the renal hilum was excluded from measurement. Importantly, all measurements were performed by readers blinded for patient number and previous TKV measurements [9]. Separate KV for left and right kidneys was determined using MATLAB software to separate the measured TKV. The Sheffield TKV Tool was used (TD) to measure KV on this dataset in a blinded manner and its performance relative to the reference manual segmentation values was then assessed.

Statistical Analyses

Baseline demographics are reported as mean ± standard deviation (SD). Agreement was assessed using Bland-Altman (BA) analyses to determine the mean difference between TKV for the various methods. Both actual and percentage (%) difference in volume were evaluated. For development of the Sheffield TKV Tool, sample size was calculated for BA agreement assessment [19] using the level of significance ρ (Type I error), a power value (Type II error), expected mean, SD and maximum acceptable percentage (%) volume difference between reference and TKV Tool measurements [19]. Based on published literature [11, 15], the expected mean (bias in BA plot) was 2%, the expected SD (precision in BA plot) 5% and the maximum allowed difference 15% (greater than mean+2*SD) [19]. Thus, for ρ of 0.05, a power of 0.80, the minimum required number of TKV measurement pairs was 60. Bias (mean) obtained from different methods was assessed using paired sample t-test.

Spatial overlap between segmentation outlines was determined using the dice similarity coefficient (DSC) [20]. A DSC value of 1 implies complete overlap while a value of 0 implies no overlap. Inter- and intra-operator variability was also assessed by coefficient of variation (CoV) [21].

Results

Characteristics of the development cohort

61 ADPKD patients (32 female, 29 male) with a mean age of 45±14 (20-77) years and eGFR of 76±32 ml/min/1.73m^2 (33-175 ml/min/1.73m^2) participated in this study. They represented a wide spectrum of disease with gold standard TKV ranging (mean±SD) between 258-3680 ml (1167±798 ml). Their kidneys had variable morphology (shape, size and heterogeneous cysts) (Figure 2) and 42 (69%) patients had liver cysts. Based on the Mayo imaging classification [6], 51 patients were categorised as Class 1 (typical) and 10 patients as Class 2 (atypical disease). Class 1 patients were further sub-divided into five prognostic groups (1A-1E) [6].

Performance of the Sheffield TKV Tool

Table 1 summarises the performance of the Sheffield TKV Tool, the ellipsoid, mid-slice and MIROS methods compared to the reference gold standard manual method in all 61 patients (122 kidneys). The mean TKV of 1153±786 ml (258-3737 ml), measured by the Sheffield TKV Tool, was close to manually measured TKV (1167±798 ml; 258-3680 ml), whereas the ellipsoid (1238±742 ml; 261-3437 ml), mid-slice (1196±827 ml; 276-4082 ml) and MIROS (1182±821 ml; 261-3780 ml) methods overestimated TKV.

In terms of volume error (Table 1, Figure 3), the Sheffield TKV Tool performed more accurately and with greater precision with a mean TKV difference of -0.3±3.8% compared to the ellipsoid (3.1±14.1%), mid-slice (3.8±9.2%) and MIROS (1.4±5.1%) methods. Paired sample t-test demonstrated that there was no statistically significant difference (ρ (2-tailed) = 0.008) between bias obtained using the Sheffield TKV Tool and MIROS method. The Sheffield TKV Tool demonstrated no particular bias ($\pm \leq 0.5\%$) and had a narrower 95% confidence interval. In contrast, the estimating equations were less reliable with a positive bias (overestimation of TKV) and more variable results (wider 95% confidence intervals).

Figure S1 shows the high agreement (0.89±0.06 (R KV) and 0.90±0.04 (L KV) of the DSC for the Sheffield TKV Tool compared to the manual method. There was high inter- (-0.5±3.5%, CoV: 2.3) and intra- (0.5±2.2%, CoV: 1.6) operator reproducibility for manual TKV measurements. Inter- (1.1±2.9%, CoV: 2.2) and intra- (-0.2±1.3%, CoV: 0.8) operator reproducibility for the Sheffield TKV Tool was higher than for manual measurements (Table 2).

Validation of the Sheffield TKV Tool in an external dataset

The MRI of 65 patients (25 female, 40 male) with ADPKD, mean age 50±8 (26-61) years and eGFR of 52±13 (33-78) ml/min/1.73m^2 with a representative spectrum of disease with gold standard TKV ranging (mean±SD) between 400-7431 ml (2408±1806 ml) were used. These patients also had more severe associated polycystic liver disease. Table 3 summarises the performance of the Sheffield TKV Tool compared to the reference gold standard manual method (Analyze Direct) in all 65 patients (130 kidneys). The mean TKV of 2344±1806 ml, measured by the Sheffield TKV Tool, was close to the manually measured TKV (2408±1806 ml).

The mean volume error (Table 3, Figure 4c) between the Sheffield TKV Tool and the manual reference for the external dataset was 3.45±3.96%. This overall positive mean (bias) difference indicates the manual volume was greater than the value obtained by the tool. This was attributed to the method of manual segmentation, which leaves a slight border around the kidney. However, standard deviation (precision) values were comparable to the initial results obtained during development of the tool (Table 1). The higher mean volume error for right KV is likely to be secondary to the low contrast between the right kidney and liver, which was often very cystic.

Time taken to measure TKV

The average time taken to measure TKV by manual segmentation was 44±18 minutes. In comparison, the Sheffield TKV Tool took 5.6±1.5 minutes on the Sheffield cohort and 5±3 minutes on the external validation cohort. The mean time to perform ellipsoid, mid-slice and MIROS methods was 4.5±0.6, 3.2±0.8 and 6.5±2.2 min respectively.

Table 4 shows the number of misclassified patients assigned to Mayo imaging classes (1A-1E) based on TKV calculated using various methods (Manual, Ellipsoid, Mid-slice, MIROS and Sheffield TKV Tool). Compared to the manual method, the Sheffield TKV Tool misclassified 2 patients from Class 1C to 1B and 1 patient from Class1A to 1B. However, these two patients were misassigned from Class 1C to 1B and Class 1A to 1B by all four methods: in this case, the value for manual HtTKV was borderline between Class 1B and 1C (age: 51 years, HtTKV: 678 ml) or between Class 1B and 1A (age: 36 years, HtTKV: 256 ml) (please refer [6] Supp Table S2). In the third patient, the Sheffield TKV Tool significantly under-segmented the kidney region due to the presence of large cysts, a current limitation (see later). Overall, class assignment based on the Sheffield TKV Tool was comparable to the Mid-Slice and MIROS

methods and performed better than the Ellipsoid method which misclassified 8 patients in total, 6 between Classes 1B and 1C (Table 4).

Discussion

We report a new semi-automated method (Sheffield TKV Tool) to measure TKV from MRI in ADPKD. The tool requires minimal user interaction to define a kidney outline from a coronal slice. Compared to manual segmentation, it performed with high accuracy in an unselected group of patients with a wide spectrum of disease represented by kidneys with highly variable morphology, cyst burden, intensity distribution and extensive range of TKV (range 258-7431ml). Importantly, it performed with high precision with no bias in measurements of right or left kidneys, high agreement (mean DSC: 0.90±0.05, TKV difference: -0.3±3.8%) and reproducibility (1.1±2.9%) compared to the manual method. Validation in a representative external group of 65 patients with ADPKD confirmed good performance (mean volume error 3.45±3.96%) with the positive bias caused by the method of manual segmentation which includes a slight border around the kidney.

A direct comparison between the Sheffield TKV Tool and two estimation methods (ellipsoid and mid-slice) in the same patients showed that it was more accurate and precise than either. Unlike the estimation methods, it also clearly outlined the kidney boundaries: these could be used as a precursor for the segmentation of renal cysts [22]. The Sheffield TKV Tool also performed as well as the MIROS method in terms of precision and accuracy (Table 1). MIROS requires more manual interaction for larger kidneys and will therefore likely require more time since the user must draw a polygon in-between slices to initiate kidney segmentation unlike the Sheffield TKV Tool where manual interaction is independent of kidney size.

The Sheffield TKV Tool also performed better (TKV difference: -0.3±3.8%) when compared to other published methods [11, 23] (Table S1). Turco et al. [23] reported a greater volume difference (-1.3±3.9%) in 30 patients with a TKV range of 693-2029 ml. Kim et al. reported a volume error of 4.2±16.8% in 30 patients for training and 30 for validation and a volume range of 177-2634 ml with their automatic method [11]. Although no manual interaction is required when measuring TKV with this technique, large volume errors (≥40% in 4 patients) resulted in the subsequent need for manual verification after segmentation.

In terms of efficiency, the reduced time required for the Sheffield TKV Tool would enable 8-10 TKV measurements to be performed in the time taken for a single manual TKV measurement. It performed particularly well on larger kidneys and in a wider range of kidney volumes (258-7431ml) than previously reported (largest 2837ml) [15, 23].

Liver cysts can cause considerable challenges when measuring TKV because of the close proximity of the liver with the right kidney and less often the left kidney (with much enlarged polycystic livers), since the distribution of cysts between the two neighbouring organs can be indistinguishable. However in most cases, the Sheffield TKV Tool was able to distinguish between liver and kidney cysts even when the visual boundaries appeared vague. There was no influence of imaging classification on the performance of the Sheffield TKV Tool: it performed equally well in Class 1 and Class 2 patients. This is a considerable advantage since no patients requiring TKV measurements need to be excluded.

The current limitations of the Sheffield TKV Tool are a slight under-segmentation and measurement of TKV in kidneys associated with exophytic cysts or over-segmentation associated with large blood vessels especially when the kidney regions are small (Figure S1j).

It has not yet been tested for serial measurments of TKV measurements to monitor natural history or response to treatment. A future goal is therefore to apply image registration techniques for this purpose [24]. Finally, the misclassification of 2 patients to a lower risk class (1C to 1B) based on TKV suggests that in cases with borderline TKV values between classes or with atypical outlines leading to under-segmentation, manual reanalysis may be required [6].

In summary, the accuracy, reproducibility and rapidity of the Sheffield TKV Tool highlights its potential for wider adoption to measure TKV as a prognostic marker routinely in patients with ADPKD.

Acknowledgements

We are grateful to our many patients for their willing and generous participation. We thank Mike Bradburn, Senior Medical Statistician, School of Health and Related Research for helpful early discussion. Initial development of the Sheffield TKV Tool was funded by the award of an INSIGNEO bursary funded by Sheffield Teaching Hospitals Charity and development by a Medical Research Council (MRC) Confidence in Concept Award. RJS was supported through a Clinical Lectureship from the National Institute for Health Research (NIHR) UK, an NIHR Clinical Trials Fellowship and an Academy of Medical Sciences UK Starter Grant.

Conflict of Interest Statement

The authors have nothing to disclose.

References

1. Schrier, R.W., *Optimal care of autosomal dominant polycystic kidney disease patients.* Nephrology (Carlton), 2006. **11**(2): p. 124-30.
2. Ong, A.C.M., Devuyst, O., Knebelmann, B., Walz, G., ERA-EDTA Working Group for Inherited Kidney Diseases., *Autosomal dominant polycystic kidney disease: the changing face of clinical management.* The Lancet, 2015. **385**(9981): p. 1993-2002.
3. Grantham, J.J., S. Mulamalla, and K.I. Swenson-Fields, *Why kidneys fail in autosomal dominant polycystic kidney disease.* Nat Rev Nephrol, 2011. **7**(10): p. 556-66.
4. Perrone, R.D., et al., *Total Kidney Volume Is a Prognostic Biomarker of Renal Function Decline and Progression to End-Stage Renal Disease in Patients With Autosomal Dominant Polycystic Kidney Disease.* Kidney Int Rep, 2017. **2**(3): p. 442-450.
5. Chapman, A.B., et al., *Kidney volume and functional outcomes in autosomal dominant polycystic kidney disease.* Clin J Am Soc Nephrol, 2012. **7**(3): p. 479-86.
6. Irazabal, M.V., Rangel, L. J., Bergstralh, E. J., et al. , *Imaging classification of autosomal dominant polycystic kidney disease: a simple model for selecting patients for clinical trials.* J Am Soc Nephrol, 2015. **26**(1): p. 160-72.
7. Torres, V.E., Chapman, A. B., Devuyst, O., *Tolvaptan in patients with autosomal dominant polycystic kidney disease.* N Engl J Med, 2012. **367**(25): p. 2407-18.
8. Gansevoort, R.T., Arici, M., Benzing, T. et al. , *Recommendations for the use of tolvaptan in autosomal dominant polycystic kidney disease: a position statement on behalf of the ERA-EDTA Working Groups on Inherited Kidney Disorders and European Renal Best Practice.* Nephrol Dial Transplant, 2016. **31**(3): p. 337-48.
9. Spithoven, E.M., van Gastel, M. D., Messchendorp, A. L., *Estimation of total kidney volume in autosomal dominant polycystic kidney disease.* Am J Kidney Dis, 2015. **66**(5): p. 792-801.
10. Zollner, F.G., Svarstad, E., Munthe-Kaas, A. Z., Schad, L. R., Lundervold, A., Rorvik, J., *Assessment of kidney volumes from MRI: acquisition and segmentation techniques.* AJR Am J Roentgenol, 2012. **199**(5): p. 1060-9.
11. Kim, Y., Ge, Y., Tao, C., et al., , *Automated Segmentation of Kidneys from MR Images in Patients with Autosomal Dominant Polycystic Kidney Disease.* Clin J Am Soc Nephrol, 2016. **11**(4): p. 576-84.
12. Sharma, K., Rupprecht, C., Caroli, A. et al.,, *Automatic Segmentation of Kidneys using Deep Learning for Total Kidney Volume Quantification in Autosomal Dominant Polycystic Kidney Disease.* Sci Rep, 2017. **7:2049**(1): p. 1-10.
13. Kline, T.L., Korfiatis, P., Edwards, M. E., et al., , *Performance of an Artificial Multi-observer Deep Neural Network for Fully Automated Segmentation of Polycystic Kidneys.* J Digit Imaging, 2017.
14. Levey, A.S., Stevens, L. A., Schmid, C. H., et al. , *A new equation to estimate glomerular filtration rate.* Ann Intern Med, 2009. **150**(9): p. 604-12.
15. Kline, T.L., Edwards, M. E., Korfiatis, P., Akkus, Z., Torres, V. E., Erickson, B. J., *Semiautomated Segmentation of Polycystic Kidneys in T2-Weighted MR Images.* AJR Am J Roentgenol, 2016. **207**(3): p. 605-13.
16. Zhang, Y., Matuszewski, B. J., Shark, L-K., Moore, Moore, C. J., *Medical Image Segmentation Using New Hybrid Level-Set Method*, in *Information Visualization in Biomedical Informatics*. 2008. p. 71-76.
17. King, B.F., Reed, J. E. , Bergstralh, E. J. , Sheedy II, P. F., Torres, V. E. , *Quantification and Longitudinal Trends of Kidney, Renal Cyst, and Renal Parenchyma Volumes in Autosomal Dominant Polycystic Kidney Disease.* J Am Soc Nephrol, 2000. **11**: p. 1505–1511.
18. Bae, K.T., Tao, C., Wang, J. et al. , *Novel Approach to Estimate Kidney and Cyst Volumes Using Mid-Slice Magnetic Resonance Images in Polycystic Kidney Disease.* American Journal of Nephrology, 2013. **38**: p. 333-341.

19. Lu, M.J., Zhong, W. H., Liu, Y. X., Miao, H. Z., Li, Y. C. Ji, M. H., *Sample Size for Assessing Agreement between Two Methods of Measurement by Bland-Altman Method.* Int J Biostat, 2016. **12**(2).
20. Zou, K.H., Warfield, S. K., Bharatha, A., et al., *Statistical Validation of Image Segmentation Quality Based on a Spatial Overlap Index.* Acad Radiol., 2004. **11**(2): p. 178–189.
21. Bland, J.M., and Altman, D. G.,, *Statistics Notes: Measurement error proportional to the mean.* British Medical Journal, 1996. **313**(7049): p. 106.
22. Bae, K., Park, B., Sun, H., et al., , *Segmentation of individual renal cysts from MR images in patients with autosomal dominant polycystic kidney disease.* Clin J Am Soc Nephrol, 2013. **8**(7): p. 1089-97.
23. Turco, D., Severi, S., Mignani, R., Aiello, V., Magistroni, R., Corsi, C., *Reliability of Total Renal Volume Computation in Polycystic Kidney Disease From Magnetic Resonance Imaging.* Acad Radiol, 2015. **22**(11): p. 1376-84.
24. Kline, T.L., Korfiatis, P. Edwards, M. E., et al., , *Automatic total kidney volume measurement on follow-up magnetic resonance images to facilitate monitoring of autosomal dominant polycystic kidney disease progression.* Nephrol Dial Transplant, 2016. **31**(2): p. 241-248.

Tables

Table 1: Accuracy and precision of different semi-automated methods of measuring or estimating KV compared to manual segmentation.

Table 2: Intra- and inter-operator variability to assess reproducibility and precision of the Sheffield TKV Tool.

Table 3: Validation: Accuracy and precision of Sheffield TKV Tool compared to manual segmentation in an external dataset.

Table 4: Number of Class1 (A-E) ADPKD (out of 51) patients [6] misclassified based on TKV measured using various TKV measurement tools. Assignment by manual TKV measurements was used as reference.

Figure Legends

Figure 1: Flow chart of Sheffield TKV Tool. 1. The tool initially loads DICOM (T2 TRUFI) series. 2. The user selects the approximate sagittal mid-slice and identifies two points to define the kidney edge which allows selection of coronal slices that contain the kidney. 3. The selected slices are pre-processed to remove motion artefact, intra- and inter- slice intensity variations. Step 3 shows MRI slices before and after pre-processing respectively. 4. The user previews the cropped image to check the entire kidney is included within the defined region and 5. Level set method is manually initialised near the kidney region boundary (white outline) only on the mid-coronal slice, 6. The final kidney region outline (white outline) is obtained based on a hybrid level set method.

Figure 2: Representative ADPKD kidney MR images used to test Sheffield TKV Tool

Figure 3: Bland–Altman analysis of different methods (Ellipsoid, Mid-Slice, MIROS and Sheffield TKV Tool) to measure TKV compared to the reference manual method. Bland-Altman plots (bold line = mean; dashed lines = 95% confidence intervals) comparing the percentage (%) volume difference of each method to the reference manual method to measure TKV in 61 patients. The Sheffield TKV Tool demonstrates higher accuracy and precision compared to all other methods.

Figure 4: Bland–Altman analysis of Sheffield TKV Tool to measure TKV compared to the reference manual method for external (Groningen) dataset.

a. Right kidney volume b. Left kidney volume c. Total kidney volume. Bland-Altman plots (bold line = mean; dashed lines = 95% confidence intervals) comparing the percentage (%) volume difference of tool to the reference manual method to measure TKV in 65 patients.

Table 1: **Accuracy and precision of different semi-automated methods of measuring or estimating KV compared to manual segmentation**.

		volume (ml) (mean ± SD)	% volume difference (mean ± SD)	Raw volume difference (mean ± SD)
Right -KV	Manual (Reference)	563 ± 400	-	-
	Ellipsoid method*****	568 ± 394	4.5 ± 19.7	11.0 ± 129.4
	Mid-slice method	568 ± 405	1.9 ± 11.1	4.1 ± 72.6
	Sheffield TKV Tool	561 ± 392	0.5 ± 5.3	-2.8 ± 25.3
Left -KV	Manual (Reference)	597 ± 417	-	-
	Ellipsoid method*****	576 ± 378	1.7 ± 17.6	-21.9 ± 162.9
	Mid-slice method	629 ± 452	6.1 ± 12.4	31.8 ± 79.0
	Sheffield TKV Tool	592 ± 419	-0.7 ± 5.5	-5.1 ± 29.7
TKV	Manual (Reference)	1167 ± 798	-	-
	Ellipsoid method*****	1238 ± 742	3.1 ± 14.1	-10.6 ± 223.76
	Mid-slice method	1196 ± 827	3.8 ± 9.2	35.9 ± 104.9
	MIROS Tool	1182 ± 821	1.4 ± 5.1	21.7 ± 60.8
	Sheffield TKV Tool	1153 ± 786	-0.3 ± 3.8	-7.9 ± 41.8

Results are shown for all 61 patients (122 kidneys). Negative values indicate under-estimation of KV compared to manual segmentation. Various methods were tested on images of kidneys with manual volumes (analyst A) as reference. The Sheffield TKV Tool was more accurate and precise compared to the other methods with no bias for either left or right kidneys.

SD = standard deviation; KV = kidney volume; TKV = total kidney volume.

*****Results for Ellipsoid method is shown only for Class1 (typical) patients (51 patients).

For MIROS, no separate volumes were obtained for left and left kidneys, thus results are reported for total kidney volume (TKV).

Table 2: Intra- and inter-operator variability to assess reproducibility and precision of the Sheffield TKV Tool.

	Comparison		% volume difference (mean ± SD)	Raw volume difference (ml) (mean ± SD)	CoV (Coefficient of Variation) (%)
Right KV	Manual segmentation	Intra-operator ($B^{manual1}$ vs. $B^{manual2}$)	1.1 ± 2.6	5.4 ± 10.6	1.9
		Inter-operator (A^{manual} vs. $B^{manual1}$)	0.9 ± 3.4	7.5 ± 12.5	2.4
	Sheffield TKV Tool	Intra-operator ($B^{SheffieldTKVTool1}$ vs. $B^{SheffieldTKVTool2}$)	-0.1 ± 1.3	1.3 ± 4.2	1.0
		Inter-operator ($A^{SheffieldTKVTool}$ vs. $B^{SheffieldTKVTool1}$)	1.5 ± 4.6	3.5 ± 22.1	3.4
Left KV	Manual segmentation	Intra-operator ($B^{manual1}$ vs. $B^{manual2}$)	0.1 ± 2.2	1.0 ± 11.0	1.5
		Inter-operator (A^{manual} vs. $B^{manual1}$)	0.2 ± 3.6	0.4 ± 14.7	2.5
	Sheffield TKV Tool	Intra-operator ($B^{SheffieldTKVTool1}$ vs. $B^{SheffieldTKVTool2}$)	-0.3 ± 1.3	-1.0 ± 6.9	0.9

		Inter-operator ($A^{SheffieldTKVTool}$ vs. $B^{SheffieldTKVTool1}$)	1.0 ± 3.1	7.4 ± 17.6	2.3

Results obtained from a subset of 20 patients (40 kidneys). KV= kidney volume; SD = standard deviation.

Table 3: Validation: Accuracy and precision of Sheffield TKV Tool compared to manual segmentation using external dataset

		volume (ml) (mean ± SD)	**% volume difference (mean ± SD)**	**Raw volume difference (mean ± SD)**
Right -KV	Manual (Reference)	1149±871	-	-
	Sheffield TKV Tool	1109±862	3.91±5.24	40.35±60.38
Left -KV	Manual (Reference)	1259±966	-	-
	Sheffield TKV Tool	1235±981	3.14±4.95	23.46±99.42
TKV	Manual (Reference)	2408±1806	-	-
	Sheffield TKV Tool	2344±1806	3.45±3.96	63.81±142.81

The reference TKV was provided by manual TKV measurements at University of Groningen using Analyze Direct 11.0 software (Spithoven Est TKV AJKD 2015). SD = standard deviation; KV = kidney volume; TKV = total kidney volume.

Table 4: Number of Class1 (A-E) ADPKD (out of 51) patients [6] misclassified based on TKV measured using various TKV measurement tools. Assignment by manual TKV measurements was used as reference.

TKV measurements methods	Class 1 misclassification					
	A to B	B to A	B to C	C to B	C to D	Total
Ellipsoid method	1	0	3	3	1	**8**
Mid-slice method	1	0	0	1	1	**3**
MIROS tool	1	1	0	1	0	**3**
Sheffield TKV Tool	1	0	0	2	0	**3**

Figure 1

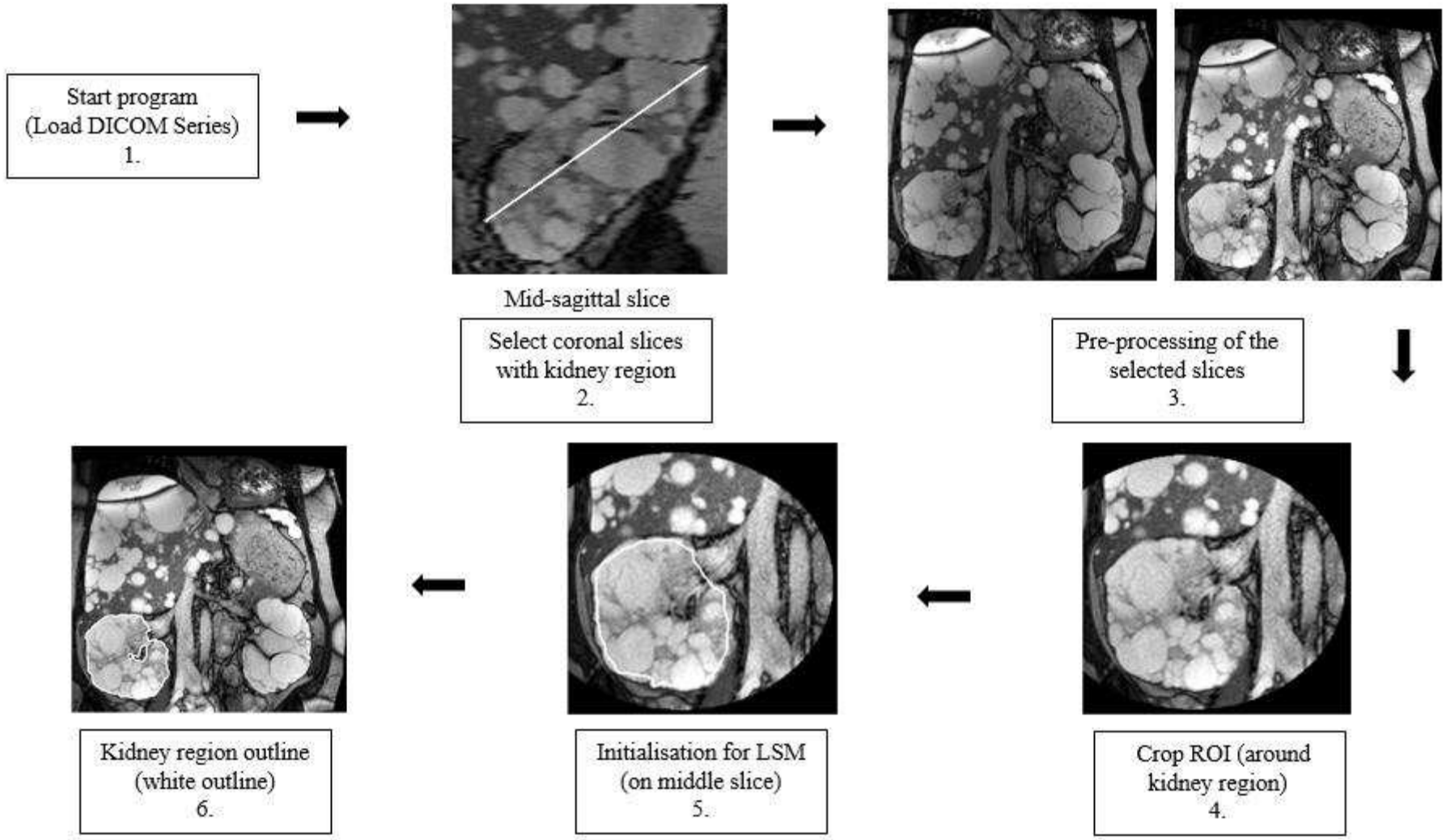
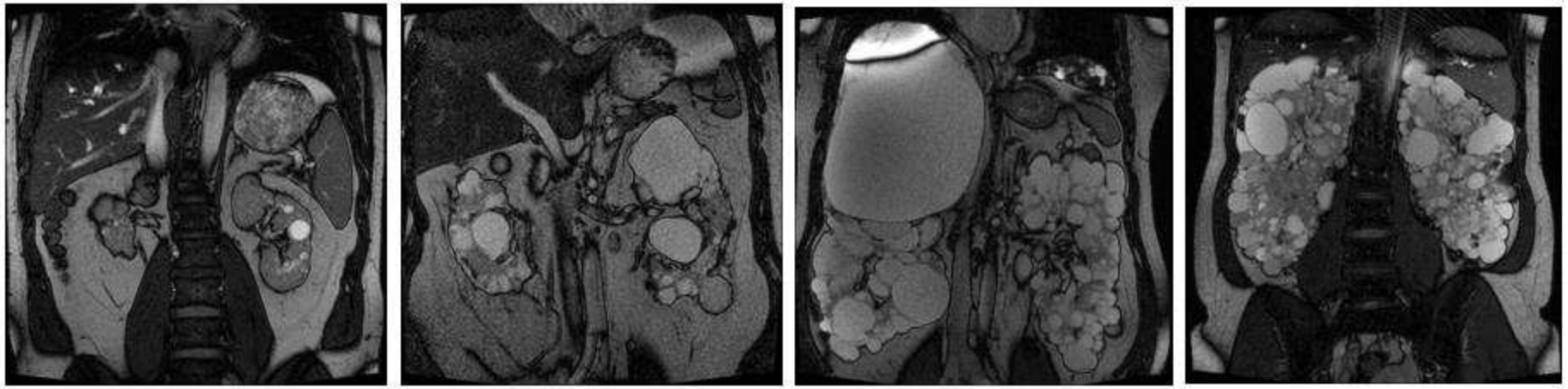
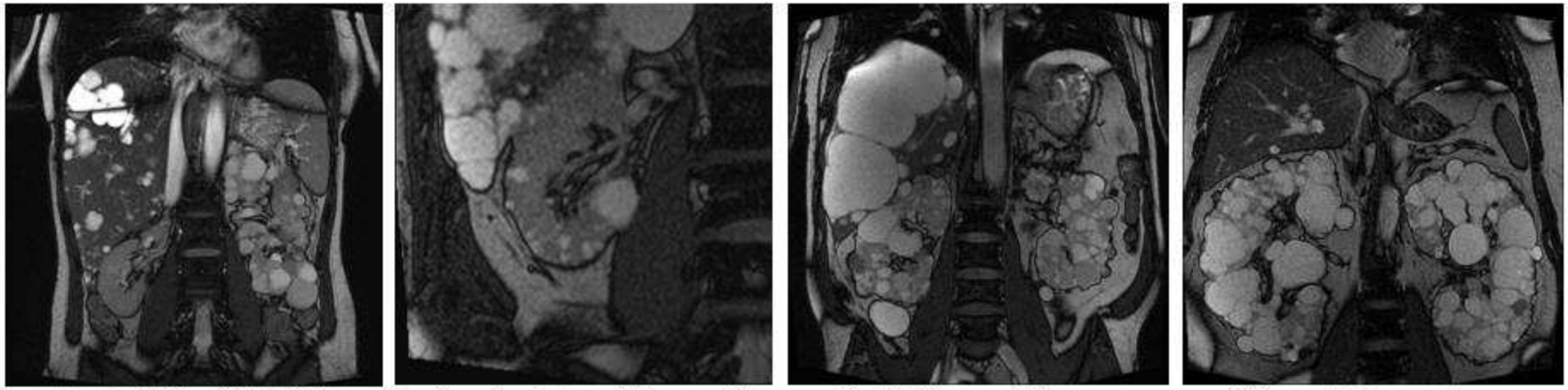

Start program
(Load DICOM Series)
1.
Mid-sagittal slice
Select coronal slices
with kidney region
2.
Pre-processing of the
selected slices
3.
Crop ROI (around
kidney region)
4.
Initialisation for LSM
(on middle slice)
5.
Kidney region outline
(white outline)
6.

Figure 2

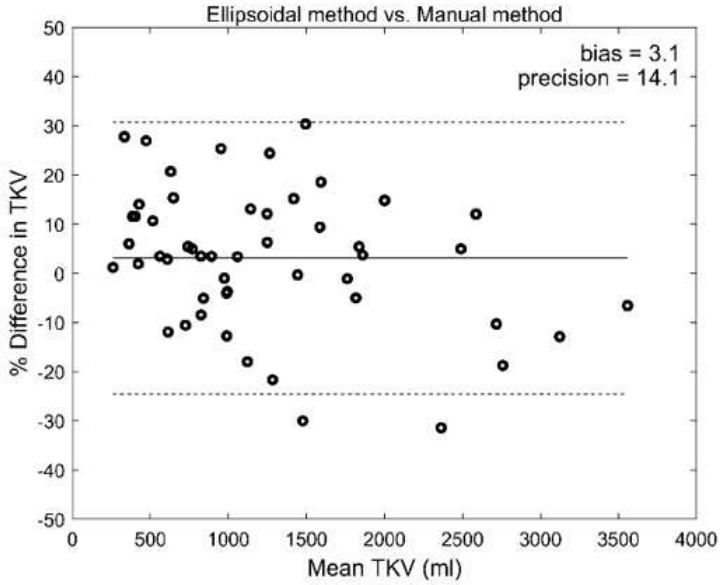


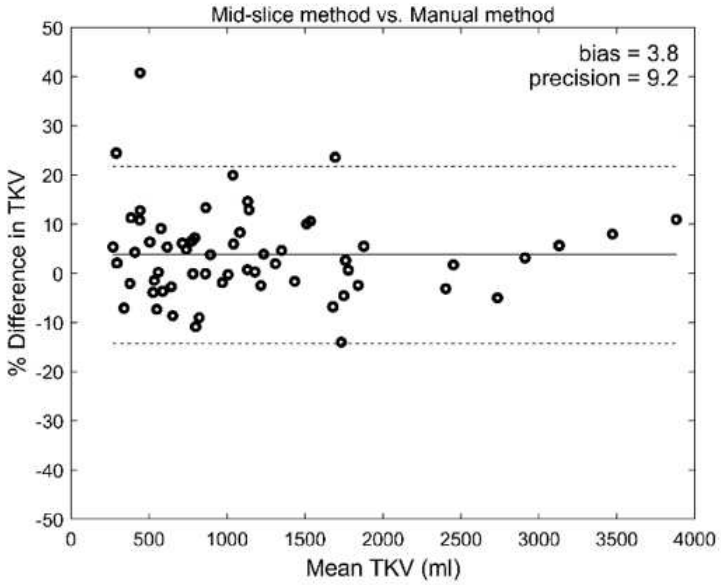
Small kidneys | Exophytic cystic kidney | Cystic kidneys and large liver cyst | Cysts with different morphology

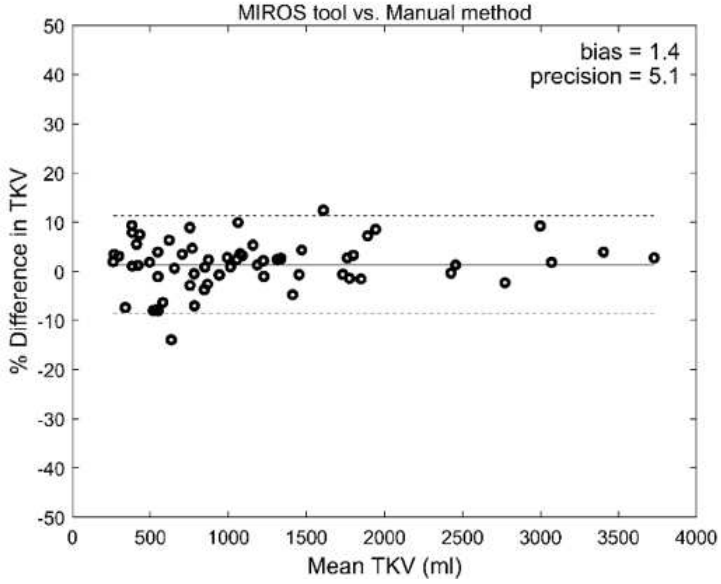


Atypical (Class 2) ADPKD | No clear edge between kidney and liver | Cystic kidney and liver | Kidney with large cysts

Figure 3


Ellipsoidal method vs. Manual method
bias = 3.1
precision = 14.1
% Difference in TKV
Mean TKV (ml)


Mid-slice method vs. Manual method
bias = 3.8
precision = 9.2
% Difference in TKV
Mean TKV (ml)


MIROS tool vs. Manual method
bias = 1.4
precision = 5.1
% Difference in TKV
Mean TKV (ml)

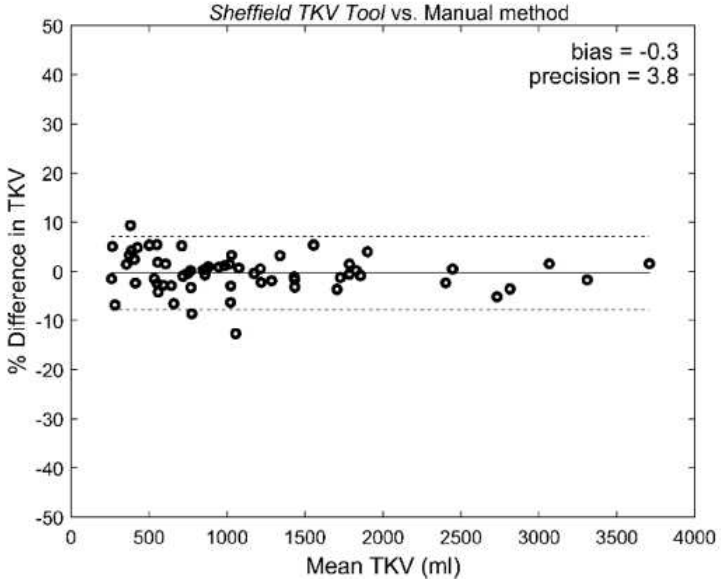
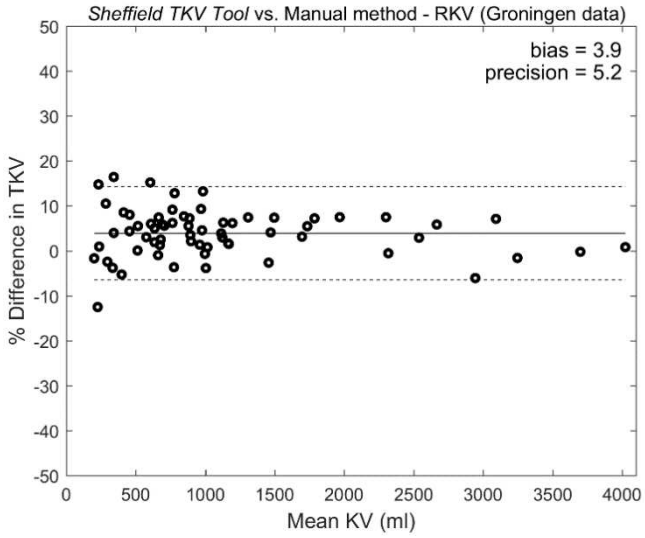
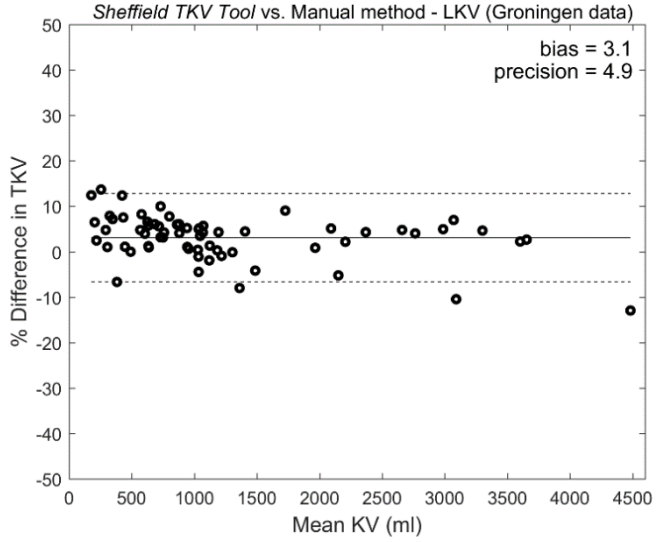

Sheffield TKV Tool vs. Manual method
bias = -0.3
precision = 3.8
% Difference in TKV
Mean TKV (ml)

Figure 4


Sheffield TKV Tool vs. Manual method - RKV (Groningen data)
bias = 3.9
precision = 5.2
% Difference in TKV
Mean KV (ml)


Sheffield TKV Tool vs. Manual method - LKV (Groningen data)
bias = 3.1
precision = 4.9
% Difference in TKV
Mean KV (ml)

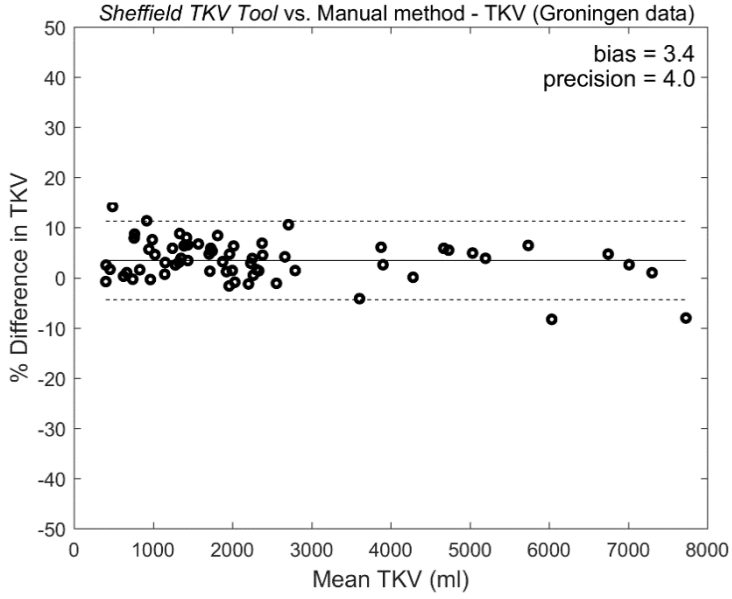
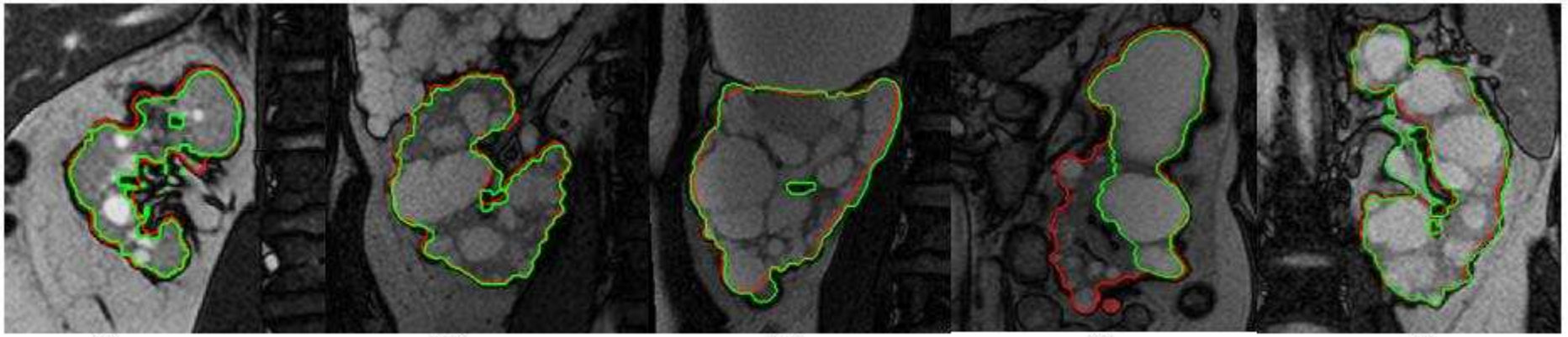
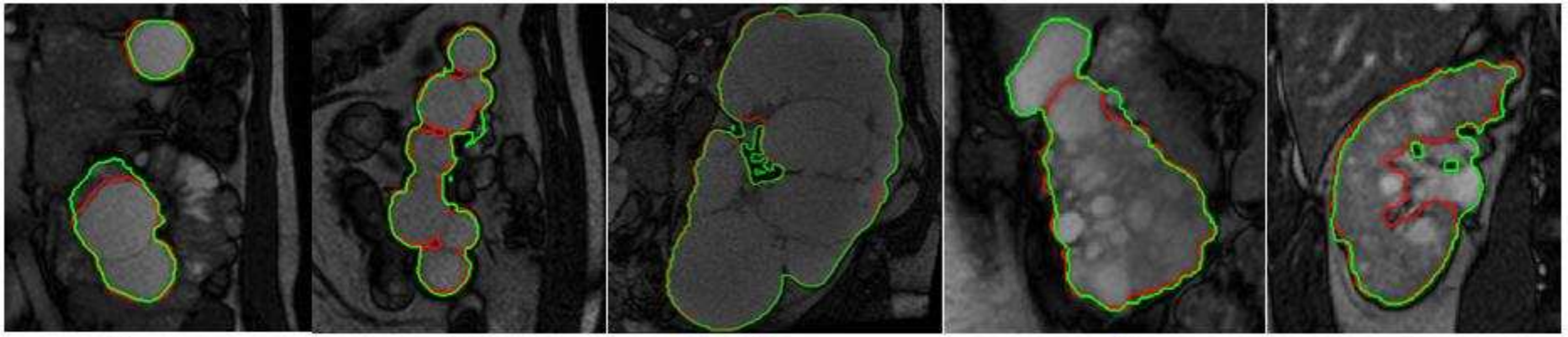

Sheffield TKV Tool vs. Manual method - TKV (Groningen data)
bias = 3.4
precision = 4.0
% Difference in TKV
Mean TKV (ml)

Figure S1



(a) (b) (c) (d) (e)

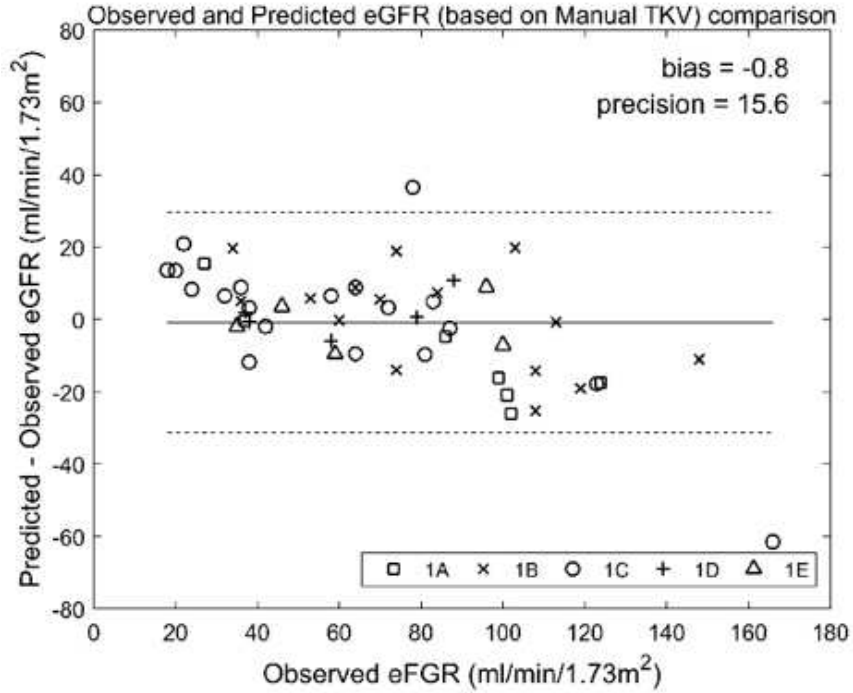
Right kidney

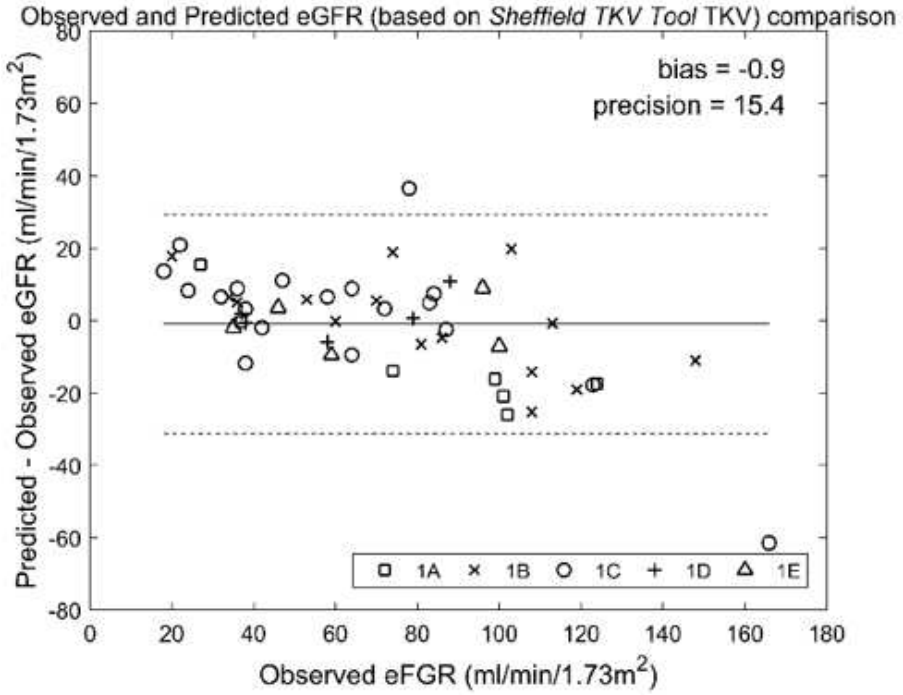


(f) (g) (h) (i) (j)

Left kidney

Figure S2


Observed and Predicted eGFR (based on Manual TKV) comparison
bias = -0.8
precision = 15.6
Predicted - Observed eGFR (ml/min/1.73m²)
Observed eFGR (ml/min/1.73m²)
1A
1B
1C
1D
1E


Observed and Predicted eGFR (based on *Sheffield TKV Tool* TKV) comparison
bias = -0.9
precision = 15.4
Predicted - Observed eGFR (ml/min/1.73m²)
Observed eFGR (ml/min/1.73m²)
1A
1B
1C
1D
1E