Original Article

Intraobserver and interobserver variations in cortical transit time measurement in children with pelviureteric junction obstruction

ABSTRACT

Cortical transit time (CTT) has recently been shown to be a useful parameter in the management of children with pelviureteric junction obstruction (PUJO). The aim of this study was to assess intraobserver and interobserver variations in the qualitative and quantitative assessment of CTT in children with PUJO. A retrospective study was performed, and ^{99m}Tc-MAG3 renogram images of all children with PUJO performed from January 2016 to December 2017 were retrieved. The images were assessed by three observers at two different time points. CTT was qualified as delayed if CTT was more than 3 min else; it was noted as normal. The intraobserver and interobserver variations in the results of the CTT of the normal kidney and affected kidney both before and after surgery were studied. The kappa statistic was used to compare the interobserver variation of qualitative interpretation of CTT. The Bland–Altman plot was used to evaluate the intraobserver and interobserver variations of the quantitative interpretation of CTT. A total of 57 ^{99m}Tc-MAG3 renal scintigraphies were evaluated. Overall, 114 renal units were evaluated with 51 normal renal units and 63 renal units with PUJO. Of these, 63 renal units with PUJO, 31 renal units had been operated upon, whereas the remaining 32 renal units had no intervention at the time of the study. The kappa statistic in the normal, affected operated, and affected unoperated kidneys was interpreted as almost perfect, substantial to almost perfect, and moderate to substantial level of agreement, respectively. The Bland–Altman plot revealed a large mean difference and wide 95% limits of agreement in affected kidneys (both operated and unoperated). The study concludes that the qualitative CTT interpretation in the affected renal unit which is most commonly used in recent studies is a reliable and reproducible parameter in children with PUJO. The quantitative measurement had wide inter- and intraobserver variation for clinical use.

Keywords: Cortical transit time, hydronephrosis, pelviureteric junction obstruction, pyeloplasty, renogram

INTRODUCTION

The evaluation of cortical transit time has recently been found useful in the management of children with pelviureteric junction obstruction (PUJO).[1,2] It has been suggested that delayed cortical transit time can predict deterioration in children with PUJO.[1-4] The method used to determine cortical transit time (CTT) is a visual assessment method.[4] This method depends on the visual assessment of the sum images produced during the ^{99m}Tc-MAG3 renogram. The CTT is considered to be delayed if there is no or almost no activity in the renal collecting system in the first 3 min. Since this parameter is a subjective assessment, hence it is prone to intraobserver and interobserver variations. Since this parameter may affect the management of patients, it is of paramount importance to

Access this article online	
Website: www.wjnm.org	Quick Response Code
DOI: 10.4103/wjnm.WJNM_37_20	

Vishesh Jain, Rakesh Kumar[1], Shamim Ahmed Shamim[1], Saurabh Arora[1], Kalaivani Mani[2], Devendra Kumar Yadav, Prabudh Goel, Anjan Dhua, Sandeep Agarwala

Departments of Pediatric Surgery, [1]Nuclear Medicine and [2]Biostatistics, All India Institute of Medical Sciences, New Delhi, India

Address for correspondence: Dr. Vishesh Jain, Department of Pediatric Surgery, All India Institute of Medical Sciences, Ansari Nagar, New Delhi - 110 029, India. E-mail: dr.vishesh79@gmail.com

Submitted: 18-May-2020, **Revised:** 16-May-2020, **Accepted:** 25-May-2020, **Published:** 14-Sep-2020

How to cite this article: Jain V, Kumar R, Shamim SA, Arora S, Mani K, Yadav DK, *et al.* Intraobserver and interobserver variations in cortical transit time measurement in children with pelviureteric junction obstruction. World J Nucl Med 2021;20:38-45.

study the agreement for the measurements of this parameter. This study was planned to assess the intraobserver reliability and interobserver agreement in the measurement of CTT on visual assessment of renograms in children with PUJO.

METHODS

A retrospective study was performed, and renograms of all children with PUJO performed from January 2016 to December 2017 were included. The study was cleared by the Institute Ethics Committee (IEC-534/03.11.2017, RP-05/2017).

Diuretic renogram

Adequate oral hydration had been advised to all children before the procedure. ^{99m}Tc-MAG3 had been prepared as per the manufacturer's instructions. Labeling quality control was checked using thin-layer chromatography. The dose of radiotracer used was 0.1 mCi/kg, with a minimum dose of 1 mCi. The scans were acquired on a Siemens E.cam, Signature series gamma camera (Siemens Medical Solutions, Illinois, USA) using a low-energy, high-resolution collimator with a parallel hole. The photopeak was centered over 140 keV with a 15%–20% window. Images were acquired with a single-head posterior view.

We followed the F+0 protocol in which an intravenous dose of furosemide (dose 1 mg/kg with a maximum dose of 40 mg) was administered at the same time as the radiopharmaceutical. The scan was acquired with the patient in the supine. The images were acquired on a 64 × 64 matrix size with a zoom of 1.23. The initial 60 frames of 15-s sum duration were used to visually assess the CTT.

Observers

Three investigators with different experience >20 years (observer 1), 11 years (observer 2), and 5 years (observer 3) in interpreting renograms reviewed the studies at two time points. The authors had a consensus meeting before the reporting of images to discuss the interpretation of CTT. The method used was that described by Piepsz *et al*.[4] The observer reported the CTT as normal or delayed. The cortical transit was considered as being delayed when no or almost no activity appeared in the subcortical structures (calyces and medulla) during at least 3 min, giving rise during these first minutes to a characteristic semilunar cortical rim. In addition, the value of CTT was also reported in seconds by reporting the frame in which the observer felt that the tracer had reached the subcortical structures.

Blinding

Only the panel of 60 images of 15-s sum duration was used. All other data such as patient name, details, prevoid images, and time-activity curve were removed from the image. The names of the patients were alphabetically arranged in MS Excel (Microsoft Corporation, Redmond, Washington, USA). A set of random numbers was produced online using a random number generator. The numbers were randomly selected from within the range of 1–57 (as 57 studies were included) without any duplication of numbers. These numbers were added in the Excel chart against the name of the patient, and the corresponding image of the patient was labeled with that number. This created a set of 57 images which were labeled with numbers from 1 to 57. Two more sets were created using the same methodology. These sets were labeled as set 1, 2, and 3. The allocation of the sets among the observer was done by opening a sealed envelope by each observer. The observer was sent the corresponding set for reporting. The images were then deleted by the observer after reporting. A new different set was again sent to each observer after a gap of 1 month for reassessment.

Statistical analysis

The intraobserver and interobserver variations in the results of the normal kidney, preoperative CTT of the affected kidney, and postoperative CTT of the affected kidney were studied. The interobserver variation in the categorical interpretation of the CTT was done by calculation Cohen's kappa statistic. The interpretation of kappa statistic was done as per the study by Landis and Kochs.[5] The intraobserver variation in the categorical interpretation of the CTT was done using McNemar's Chi-square test. The intraobserver and interobserver variations in the absolute value of CTT were done using Bland–Altman plot. The results were reported as mean difference (95% limits of agreement). Paired *t*-test was used to calculate the *P* value of the quantitative assessment of interobserver and intraobserver variations. All the analysis was done using Stata 12.0 (StataCorp Lp, College Station, Texas, USA).

RESULTS

A total of 57 ^{99m}Tc-MAG3 renal scintigraphies were evaluated by all three investigators. Of these 57 studies, 29 were preoperative studies, and the remaining 28 were postoperative studies. Three patients with six studies (three preoperative and three postoperative studies) had bilateral PUJO. There were no patients with a solitary kidney. Overall, 114 renal units were evaluated with 51 normal renal units and 63 renal units with PUJO. Of these 63 renal units with PUJO, 31 renal units had been operated upon, whereas the remaining 32 renal units had no intervention at the time of the renogram.

Normal renal units

Comparing interpretation of cortical transit time as normal or delayed

Of the 51 normal renal units, the cortical transit time was interpreted to be normal in all by observer 1 and in 50/51 (98%) by observers 2 and 3. The interobserver agreement was 96.15% with a kappa statistic of 1.0.

Comparing the absolute value of cortical transit time

The mean and standard deviation of the CTT value of normal renal units interpreted by all three observers at the two time points are mentioned in Tables 1 and 2. The Pearson correlation coefficient showed a positive correlation although small in between the observations [Tables 1 and 2].

The Bland–Altman plot of intraobserver and interobserver values of CTT in normal renal units is shown in Figure 1. The graphs show a low mean difference in both intraobserver and interobserver comparison. The mean difference ranges from 1.4 s to 11.4 s only. The 95% limits of agreement are also not very wide (<±60 s from the mean difference) suggesting a good agreement for the values of CTT in a normal kidney.

Renal units with pelviureteric junction obstruction

Before surgery

Thirty-two renal units were assessed in this subgroup.

Comparing interpretation of cortical transit time as normal or delayed

Intraobserver variation: The asymptotic symmetry test was not significant in observers 1 and 2 (with $P = 0.56$ in both). The test was significant in observer 3, as in 4 renal units, the CTT was observed as delayed in first observation but as normal in second observation (P value of 0.046).

Interobserver variation: Cohen's kappa statistic and the strength of agreement of CTT between the three observers are mentioned in Table 3.

Comparing the absolute value of cortical transit time

The mean and standard deviation of the CTT value of affected nonoperated renal units interpreted by all three observers at the two time points are mentioned in Tables 1 and 2. There was a high positive correlation in between the intraobserver and moderate positive correlation in between the interobserver comparison [Tables 1 and 2]. The Bland–Altman plot of intraobserver and interobserver values of CTT in the affected nonoperated renal unit is shown in Figure 2. On comparing the intraobserver observations [Figure 2a-c], although the mean difference is low (−4.0 s to 6.7 s) and most of the values lie within 95% limits of agreement, the limits are very wide for clinical utility (more than ± 80 s from the mean difference). Similarly, on the interobserver observations [Figure 2d-f], the mean difference is higher (12–32 s) and the 95% limits of agreement are also wide (more than ± 110 s from the mean difference).

After surgery

Thirty-one renal units were assessed in this subgroup.

Comparing interpretation of cortical transit time as normal or delayed

Intraobserver variation: The asymptotic symmetry test was not significant ($P = 1.00$) in either of the observers.

Interobserver variation: Cohen's kappa statistic and the strength of agreement of CTT between the three observers are mentioned in Table 4.

Comparing the absolute value of cortical transit time

The mean and standard deviation of the CTT value of affected operated renal units interpreted by all the three observers at the two time points are mentioned in Tables 1 and 2. There was a high positive correlation in between the observations

Table 1: The value of cortical transit time in seconds (mean±standard deviation) of the different observers at different time points

Observer	Time point	Normal (mean±SD)	*P*, *r*	Affected preoperative (mean±SD)	*P*, *r*	Affected postoperative (mean±SD)	*P*, *r*
1	1	140.3±19.6	0.49, 0.36	266.6±96.4	0.35, 0.80	220±96.5	0.75, 0.79
1	2	137.8±25.0		275.6±103.9		223.1±95.6	
2	1	145.6±25.85	0.09, 0.51	253.3±103.1	0.03, 0.83	207.7±82	0.86, 0.83
2	2	138.8±30.6		233.6±89.4		209.1±84	
3	1	134.1±29.45	0.62, 0.70	221.6±95.3	0.68, 0.89	189.1±68	0.84, 0.94
3	2	135.5±21.83		218.7±94.0		182.7±70.4	

Paired *t*-test was used to calculate *P*. *r*: Pearson correlation coefficient; SD: Standard deviation

Table 2: The *P* value (paired *t*-test) and *r* (Pearson correlation coefficient) of the interobserver comparison for the quantitative value of the cortical transit time

Observer	Time point	Normal (*P*, *r*)	Affected preoperative (*P*, *r*)	Affected postoperative (*P*, *r*)
1 and 2	1	0.12, 0.46	0.09, 0.68	0.27, 0.74
1 and 3	1	0.14, 0.32	0.02, 0.70	0.003, 0.81
2 and 3	1	0.009, 0.65	0.0001, 0.64	0.007, 0.88

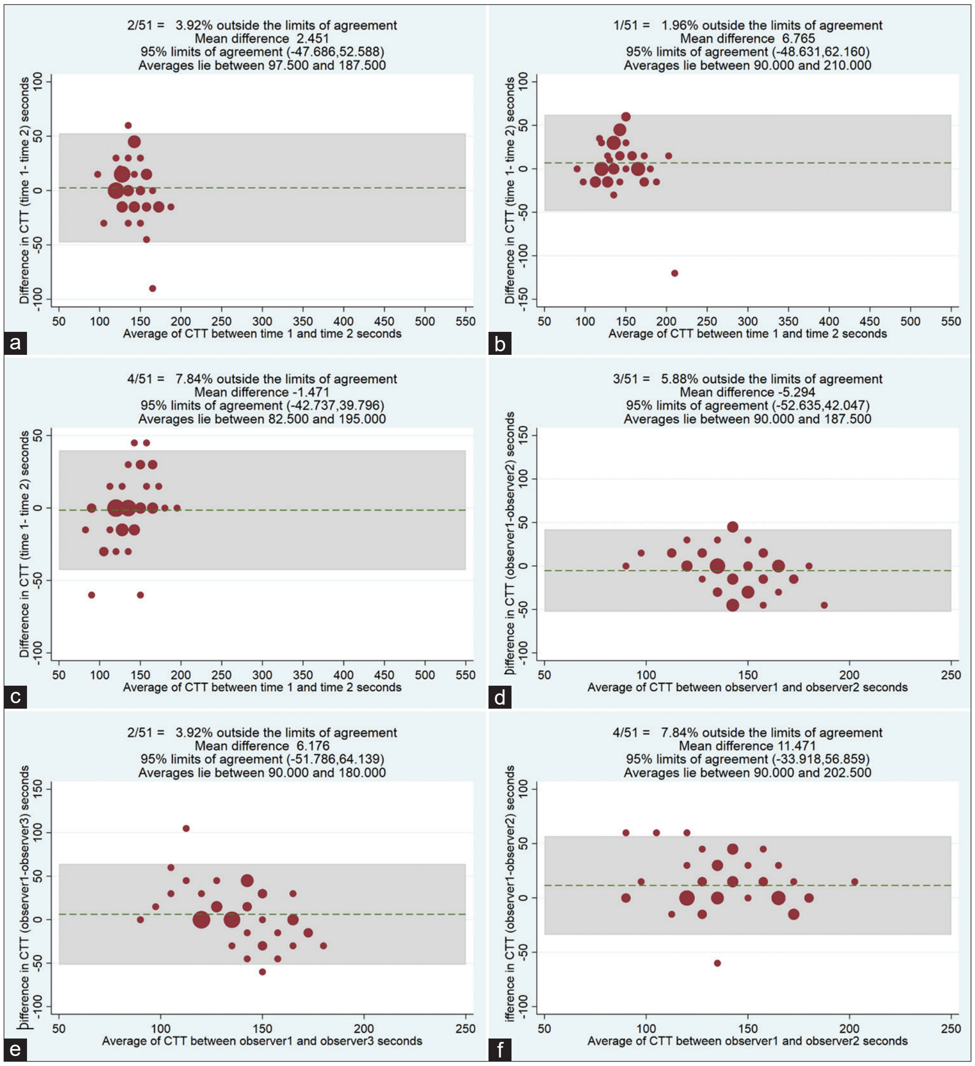

Figure 1: (a-c) Bland–Altman plot of intraobserver comparison of cortical transit time in normal kidneys. (d-f) Bland–Altman plot of interobserver comparison of cortical transit time in normal kidneys. All the plot (a-f) shows a very low mean difference (<11 s) with a narrow 95% limits of agreement

Table 3: Kappa value and the degree of agreement in the interobserver interpretation of cortical transit time as normal or delayed in affected kidneys before surgery

Observer	Kappa	Degree of agreement
Observer 1 versus 2	0.76	Substantial
Observer 1 versus 3	0.61	Substantial
Observer 2 versus 3	0.81	Almost perfect

Table 4: Kappa value and the degree of agreement in the interobserver interpretation of cortical transit time as normal or delayed in affected kidneys after surgery

Observer	Kappa	Strength of agreement
Observer 1 versus 2	0.41	Moderate
Observer 1 versus 3	0.62	Substantial
Observer 2 versus 3	0.55	Moderate

both intraobserver and interobserver [Tables 1 and 2]. The Bland–Altman plot of intraobserver and interobserver values of CTT in a normal renal unit is shown in Figure 3.

On comparing the intraobserver observations [Figure 3a-c], although the mean difference is low (3–20 s) and most of the values lie within 95% limits of agreement, the limits are very wide for clinical utility (more than ±100 s from the mean difference). On analyzing the interobserver observations [Figure 3d-f], the mean difference is high (24–57 s) and the 95% limits of agreement remain wide for clinical utility [as wide as −127, 193 s in Figure 3f].

DISCUSSION

Children with antenatally diagnosed PUJO with preserved function are managed conservatively in the postnatal period. Surgery is usually indicated if these is a fall in the differential function or increase in hydronephrosis of the affected kidney on follow up. It can be argued that the waiting for fall in function may result in irreversible loss of renal function. Clinicians and researchers are still looking for a marker or a

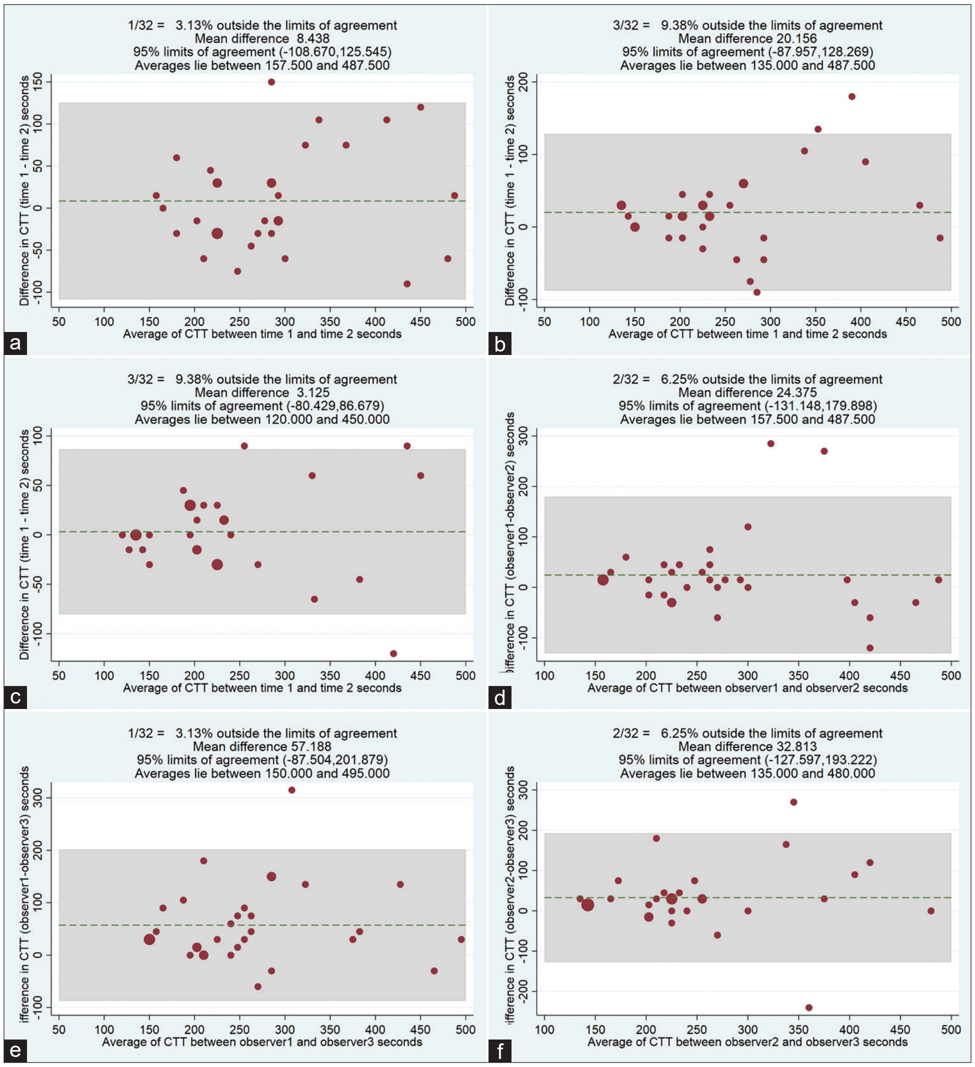

Figure 2: (a-c) Bland–Altman plot of intraobserver comparison of cortical transit time in kidneys with pelviureteric junction obstruction before surgery shows a low mean difference (3–20 s). The 95% limits of agreement are wide (>±80 s). (d-f) Bland–Altman plot of interobserver comparison of cortical transit time in kidneys with pelviureteric junction obstruction before surgery shows high mean difference (24–57 s). The 95% limits of agreement are also wide (>±110 s)

test that can predict which kidney may require surgery before there is a fall in function.

Recently, CTT has been studied retrospectively in children with PUJO and correlated with the need for surgery. These studies concluded that cortical transit time was delayed in patients who needed surgery.[2-5] CTT can be measured by the deconvolution method, but these methods are difficult to apply in children with hydronephrosis as the prerequisite of a stationary and linear system is not met.[6] A visual method was proposed to ascertain if the CTT is normal or delayed.[2,4] This method relies on analyzing summed images of the renogram for the appearance of activity in the subcortical area, i.e., calyces. This is a subjective method and hence is prone to intraobserver and interobserver variations. Before accepting CTT as a parameter to guide management in children with PUJO, it is necessary to study these variations.

Most studies have used the visual method to categorize CTT as a quantitative variable and interpreted it as normal or delayed. In addition to analyzing CTT as a qualitative variable, we also documented and compared the quantitative values reported by observers. We also did a subgroup analysis of normal, affected renal unit before surgery, and affected renal unit after surgery. This was done as we thought that variation in these subgroups would be different.

The CTT of the normal renal unit showed very less intraobserver and interobserver variations when analyzed as a qualitative variable. When assessed as a qualitative variable, the kappa statistic was 1.0 (almost perfect strength of agreement). The Bland–Altman plot also showed a low mean difference with narrow 95% limits of agreement. However, the correlation was low positive to moderate positive in between the sets of observation both intraobserver and interobserver.

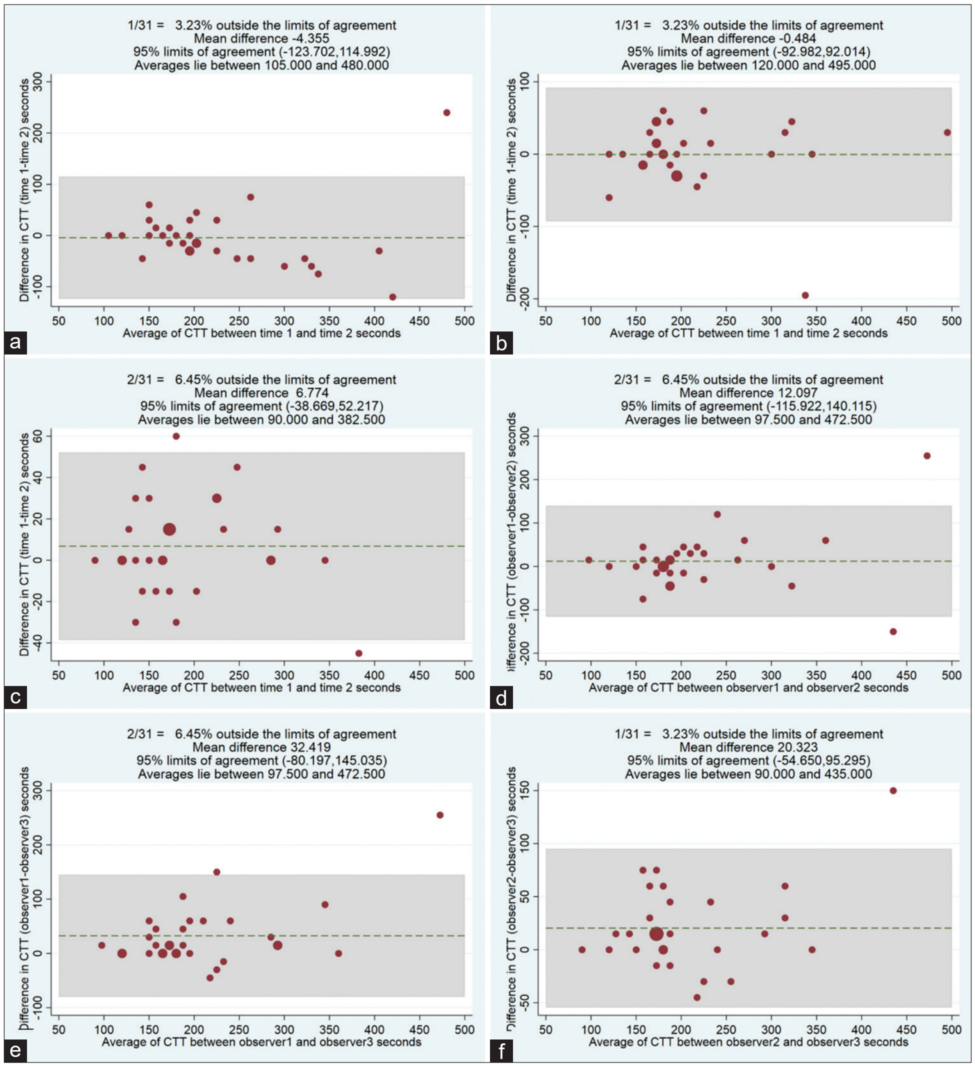

Figure 3: (a-c) Bland–Altman plot of intraobserver comparison of cortical transit time in kidneys with pelviureteric junction obstruction after surgery. The mean difference is low (−4 to 6 s) but the 95% limits of agreement are very wide for clinical utility (>±60 s). (d-f) Bland–Altman plot of interobserver comparison of cortical transit time in kidneys with pelviureteric junction obstruction after surgery. The mean difference is high (12–32 s) and the 95% limits of agreement remain wide

Analysis of qualitative measurement of CTT in affected kidneys before surgery showed high kappa statistic (substantial to almost perfect) suggestive of good interobserver reliability. McHugh recommended that a more stringent interpretation of kappa statistic as in health-care major decision is based on these observations. Hence, the level of agreement should be high.[7] The level of agreement is moderate to strong as per the value of kappa suggested by McHugh. Intraobserver reliability in one of the observers showed a significant difference. The reason for this difference is not obvious though this observer had the least experience in reporting renograms as compared to the other two observers. When assessed as a quantitative variable, the Bland–Altman plot showed a large mean difference with a wide 95% limit of agreement. In a majority, the difference was also statistically significant with a moderate to high positive correlation. This finding was noticed on both intraobserver and interobserver analyses. This shows that the affected kidneys before surgery the CTT can be reliably assessed have a qualitative variable (normal or delayed). However, due to wide variation, the quantitative value of the CTT is unreliable and unreproducible.

Analysis of qualitative measurement of CTT in affected kidneys before surgery showed kappa statistic suggestive of moderate to substantial intraobserver agreement. The interpretation as per the values recommended by McHugh is weak to moderate.[7] The symmetry test applied for testing of interobserver reliability of the qualitative measurement of CTT showed no significant difference. However, on analyzing the data, we noticed that the observer with least experience (observer 3) had interpreted 25% of the renal units differently when analyzing the same images at different time points. Since the discordance was of equal magnitude in both directions, the result in the symmetry test was not significant.

When assessed as a quantitative variable, the Bland–Altman plot showed a large mean difference with a wide 95% limit of agreement. This finding was noticed on both intraobserver and interobserver analyses. The difference between the qualitative values was not significant on intraobserver assessment, and there was high positive correlation. However, the difference was significant on interobserver assessment, and the correlation was high positive. Thus, the qualitative assessment of CTT in affected operated (normal and delayed) can be made but with less certainty as compared to affected nonoperated kidneys and probably more experience is needed in interpretation of operated kidneys. Furthermore, due to wide variation, the quantitative value of the CTT in affected operated kidneys is unreliable and unreproducible.

Most of the recent studies have used CTT as a qualitative variable in the affected kidney before surgery. Our study shows that the measurement of CTT in this manner is reliable and reproducible. The use of CTT as a qualitative variable in the affected kidney after surgery is not reliable and has weak to moderate level of agreement as per recommendations by McHugh.[7] The quantitative value of CTT in the affected kidney both before and after surgery is not reliable and has a very wide 95% limits of agreement ($> \pm 100$ s from the mean difference). This means that on quantitative assessment, the difference in values between two observers can be as high as 3–4 min. This variation in the interobserver and intraobserver quantitative measurement of CTT time in obstructed kidney limits its use in decision making. However, CTT can be reliably assessed in normal renal units both qualitatively and quantitatively.

Harper *et al*. had mentioned that CTT was assessed twice by the observer to verify intraobserver reliability.[3] The authors do not mention the presence and level of reliability in assessment by the same observer at two different time points in their study. In the study by Piepsz *et al*., the CTT was assessed both qualitatively and quantitatively by three observers.[4] The authors had mentioned that there was good agreement between the observers but do not mention the statistical method used to make this statement. They had used 60 s summed images for interpretation of CTT, whereas we had used 15 s summed images. This may account partially for more heterogeneity noticed in our study.

Another variation that is often seen in various studies reporting the cortical transit time is the time at which furosemide is injected during the renogram. Harper *et al*.[3] used a F + 20 protocol, whereas Piepsz *et al.*[4] and the index study had used a F0 protocol. We feel that for CTT calculation, the timing of administration of furosemide is not very important. The onset of action of intravenous furosemide is after 5 min and peaks at 20 min. After 5 min of injection, the decision of whether the CTT is normal or delayed is already made. However, if quantitative value of CTT is being used, then the result between quantitative values of CTT if more than 300 s may vary between the both protocols. In the current study, it was of little value as the aim was not to determine the normal values of CTT but to compare the similar study between observers.

The study also underscores the importance of experience in interpreting the renogram for calculation of CTT as it is a subjective interpretation. This is evident in our study as the difference in the values of CTT between the two more experienced observers was not statistically significant but was significant in almost all comparison when the less experienced observer was involved. However, since the intraobserver changes for the latter were not significant, we feel that a better and more comprehensive meeting may help to mitigate the large variations observed.

The strength of this study is that it is the only study to date to assess methodically the interobserver and intraobserver variations in CTT interpretation in children with PUJO, strict protocols were used to blind the observers, and observers with variable experience were included so assess wide application of this parameter.

CONCLUSION

The study concludes that the qualitative CTT interpretation in the affected renal unit, that is, most commonly used in recent studies, is reliable and reproducible. However, the use of quantitative value of CTT in obstructed kidneys cannot be used reliably to assess the severity of obstruction or improvement after surgery due to marked intraobserver and interobserver variations.

Financial support and sponsorship

The study was funded by the intramural research project grant, Organization: All India Institute of Medical Sciences, Place: New Delhi.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

1. Schlotmann A, Clorius JH, Rohrschneider WK, Clorius SN, Amelung F, Becker K. Diuretic renography in hydronephrosis: Delayed tissue tracer transit accompanies both functional decline and tissue reorganization. J Nucl Med 2008;49:1196-203.
2. Schlotmann A, Clorius JH, Clorius SN. Diuretic renography in hydronephrosis: Renal tissue tracer transit predicts functional course and

thereby need for surgery. Eur J Nucl Med Mol Imaging 2009;36:1665-73.
3. Harper L, Bourquard D, Grosos C, Abbo O, Ferdynus C, Michel JL, *et al*. Cortical transit time as a predictive marker of the need for surgery in children with pelvi-ureteric junction stenosis: Preliminary study. J Pediatr Urol 2013;9:1054-8.
4. Piepsz A, Tondeur M, Nogarède C, Collier F, Ismaili K, Hall M, *et al*. Can severely impaired cortical transit predict which children with pelvi-ureteric junction stenosis detected antenatally might benefit from pyeloplasty? Nucl Med Commun 2011;32:199-205.
5. Lee JN, Kang JK, Jeong SY, Lee SM, Cho MH, Ha YS, *et al*. Predictive value of cortical transit time on MAG3 for surgery in antenatally detected unilateral hydronephrosis caused by ureteropelvic junction stenosis. J Pediatr Urol 2018;14:55.
6. Piepsz A, Sixt R, Gordon I. Performing renography in children with antenatally detected pelvi-ureteric junction stenosis: Errors, pitfalls, controversies. Q J Nucl Med Mol Imaging 2010;54:350-62.
7. McHugh ML. Interrater reliability: The kappa statistic. Biochem Med (Zagreb) 2012;22:276-82.