A step-by-step regressed pediatric kidney depth formula validated by a reasonable index

Si Hongwei, MD, PhD, Chen Yingmao, PhD, Li Li, MD, Ma Guangyu, MD, Shen Liuhai, BMA, BMB, C, Wu Zhifang, MD, Shao Mingzhe, MD, Li Sijin, MD, PhD

Abstract

In predicting pediatric kidney depth, we are especially interested in that the errors of most estimates are within a narrow range. Therefore, this study was intended to use the proportion of estimates within a range of –5 to 5 mm (P5 mm) to evaluate the formulas and tried to regress a kidney depth formula for children. The enrolled children aged from 1 to 19 years were randomly sampled into group A and group B (75% and 25% of all recruits, respectively). Using data of the group A, the test formula was regressed by nonlinear regression and subsequently Passing & Bablok regression, and validated in group B. The Raynaud, Gordon, Tonnesen, Taylor, and the test formulas were evaluated in the 2 groups. Accuracy was evaluated by bias, absolute bias, and P5 mm; and precision was evaluated by correlation coefficient. In addition, root-mean square error was used as a mixed index for both accuracy and precision. Body weight, height, and age did not have significant differences between the 2 groups. In the nonlinear regression, coefficients of the formula (kidney depth = a × weight/height + b × age) from group A were in narrower 95% confidence intervals. After the Passing & Bablok regression, biases of left and right kidney estimates were significantly decreased. In the evaluation of formulas, the test formula was obviously better than other formulas mentioned above, and P5 mm for left and right kidneys was about 60%. Among children younger than 10 years, P5 mm was even more than 70% for left and right kidney depths. To predict pediatric kidney depth, accuracy and precision of a step-by-step regressed formula were better than the 4 “standard” formulas.

Abbreviations: 95% CI = 95% confidence interval, BMI = body mass index, GFR = glomerular filtration rate, OR = odds ratio, P30 = proportion of estimates within 30% deviation of measurement, P5 m = proportion of estimates within a range of –5 to 5 m, RMSE = root-mean square error.

Keywords: accuracy, kidney depth, Passing & Bablok regression, precision

1. Introduction

In calculating renography-based differential function, it may not be necessary to correct attenuation effects caused by kidney depth.[11] However, in measuring renography-based glomerular filtration rate (GFR), the depth needs to be corrected even in children.[2,3] It is because that assuming a linear attenuation coefficient of 0.153 cm–1 for 99m-technetium, 1 cm variation in depth will result in 14% change in the measurement.[4] Besides this, the depth is also important for percutaneous native kidney biopsy.[5,6] Above all, there is a need to accurately estimate the pediatric kidney depth.

Currently, kidney depth is usually estimated by equations from body weight, height, and/or age, for example, the Raynaud, Gordon, Tonnesen, and Taylor formulas (Table 1).[7–10] The first 2 equations were regressed from pediatric populations, and the first 3 were provided by the Xeleris workstation.[11] The workstation sets the Gordon formula as the default for predicting kidney depth of children (age less than 19 years), as its estimates correlated well with the measured values on lateral DMSA images (correlation coefficient, r = 0.94).[11] Besides the precision index r, other studies utilized the accuracy indices of absolute difference and mean difference in evaluating the performance of kidney depth formulas.[11] However, seldom did studies simultaneously utilize both accuracy and precision indices in the evaluation.

In the fields of science, engineering, and statistics, accuracy of a measurement system is the degree of closeness to its true value, and precision is the degree to which repeated measurements under unchanged conditions show the same results.[12] A higher precision index does not always mean a better accuracy, and vice versa. Therefore, it is important to simultaneously evaluate both accuracy and precision of a measurement system, for example, kidney depth formulas.

In addition, in evaluating kidney depth formulas, we are particularly interested in the fact that the error of most estimates is within a narrow range. Inspired by the studies on the evaluation of formulas for estimating GFR, the proportion of estimates within 30% deviation of measurement (P30) as a reasonable
accurate index,

we introduce proportion of estimates within a
range of \(-5 \text{ to } 5 \text{ mm} (P5 \text{ mm})\) to reasonably assess the accuracy of
kidney depth formulas.

Above all, this study was intended to reasonably evaluate the accuracy of kidney depth formulas and tried to regress a pediatric kidney depth formula with higher accuracy and higher precision.

2. Methods

2.1. Enrollment criteria and measurement

Data were retrospectively enrolled from 2 independent medical centers. The enrollment criterion was age <19 years. The recruits with a history of renal diseases were excluded, especially hydronephrosis and other diseases having the possibility of modifying kidney depth. Age was documented in years. Body weight (kg) and height (cm) were measured before the examinations and were accurate in integer. The protocol of the retrospective study was approved by the ethics committee at the 2 medical centers.

At the Chinese PLA General Hospital, images were acquired by the Biograph Truepoint 16 PET/computed tomography (CT) scanner (Siemens Medical Solutions, Knoxville, TN). Only the CT images (130 kV, 120 mA, and a pitch of 1 mm) were used to determine renal depth of children. At the First Hospital of Shanxi Medical University, the GE LightSpeed 64 scanner (GE Healthcare, Waukesha, WI) was used to acquire images. The standard abdomen protocol was used (120 kV, 320 mA, and a pitch of 0.984 mm). At both medical centers, images with an interval of 5 mm were reconstructed by the iteration method.

Kidney depth was measured according to the method of Ma et al.\(^1\) In brief, on transverse CT images with maximal left or right renal area, the vertical distances from back skin to the anterior and to the posterior renal boundary were measured and averaged as the kidney depth. The depths for left and right kidney were considered as the reference value and were accurate in 0.1 cm.

2.2. Statistical analysis

Using SPSS statistical software, 25% of all recruits were randomly sampled into group B and the remaining into group A. Data from group A were used to regress formulas and validated in group B. Accuracy and precision of kidney depth formulas (including the test formula and the 4 “standard” formulas in Table 1) were compared between the groups.

Using nonlinear regression procedure, the reference values were predicted from body weight, height, and/or age (or different combinations of these the physical variables). If the linear regression line between estimates and reference values did not pass through the origin point with a slope of 1, an additional Passing & Bablok regression was used to correct the estimates. The advantage of Passing & Bablok linear regression is that it does not assume normal distribution and is robust against outliers.\(^1\) In this study, the Passing & Bablok regression could make the estimates closer to the measured values and make the regression line pass through the origin point with a slope of 1.\(^1\)

Accuracy was assessed by P5 mm, mean and absolute differences (bias and absolute bias), and precision was assessed by correlation coefficient. The difference between bias and absolute bias could be illustrated by the following example. If the differences between predicted and actual values were 2 and \(-2 \text{ mm}\), bias (mean difference) and absolute bias (absolute difference) were 0 and 2 mm, respectively. Root-mean square error (RMSE) was used as a mixed index for both accuracy and precision. The 95% confidence interval (95% CI) for continuous variables was calculated by the bootstrap method.\(^1\) In the evaluation, P5 mm and correlation coefficient increased with an improvement, while bias, absolute bias, and RMSE decreased.

To explore the performance of the test formula, binary logistic regression tests were used against P5 mm range (in and out of the range from \(-5 \text{ to } 5 \text{ mm})\). In the test group, age, body weight, height, body mass index (BMI), and measured kidney depth were independent variables. The 95% CIs of odds ratio (OR) excluding 1 and Wald \(P < 0.05\) were considered as statistical significance. Discriminatory ability of the regression was tested by the \(c\) statistic. The \(c\) statistic, that is, concordance statistic, measures the predictive accuracy of logistic regression model. Its 95% CI was calculated using 5000 bootstrap replications.

Data were analyzed by SPSS statistical software (version 10.01; IBM, Chicago, IL) and MedCalc package (trial version, v 12.0; Mariakerke, Belgium). A 2-sided \(P\) value less than 0.05 was considered to be a significant level.

3. Results

Recruit characteristics are listed in Table 2. Body weight, height, and age did not have difference between the groups. About 75% children were aged between 7.75 to 18 (median 14.5) and 7 to 16.5 years (median 13) in the groups A and B, respectively. Age distribution is plotted in Fig. 1A. One sample Kolmogorov–Smirnov test indicated that age was not normally distributed (\(P = 0.000\)).

3.1. Regression of kidney depth formulas

The relations between kidney depth and patient characteristics are plotted in Fig. 2, which indicate that age, body weight, and height correlated with both left and right kidney depths, especially the ratio of weight/height. After trying some models (not presented), coefficients of Eq. (1) from group A were within narrower 95% CIs (Fig. 3), which overlapped with those from group B and all recruits. In addition, Fig. 3 also indicates that the
Table 2
Recruit characteristics.

|         | FHSM | CPGH | Group A | Group B | All patients |
|---------|------|------|---------|---------|--------------|
| n       | 122  | 93   | 162     | 53      | 215          |
| Gender (M/F) | 73/49 | 65/28 | 102/60  | 36/17   | 138/77       |
| Age, y  | 10.3±5.7 | 15.6±3.5 | 12.8±5.6 | 12.9±5.3 | 12.6±5.5     |
| Height, cm | 139.8±30.6 | 162.7±16.1 | 150.4±27.9 | 147.6±27.6 | 140.7±27.7 |
| Weight, kg | 38.4±21.6 | 53.1±17.1 | 45.2±21.3 | 43.5±20.3 | 44.8±21.0    |
| Left kidney depth, cm | 4.9±1.5 | 5.4±1.3 | 5.2±1.5 | 5.0±1.4 | 5.1±1.5    |
| Right kidney depth, cm | 5.2±1.7 | 5.5±1.4 | 5.4±1.6 | 5.2±1.5 | 5.3±1.6 |

CPGH=Chinese PLA General Hospital. FHSM=The First Hospital of Shanxi Medical University.

*P<0.001 significance between the centers or between the groups.

Kidney depth = a × W/H + b × A.

where a and b are coefficients, and W, H, and A are body weight, height, and age, respectively.

In group A, the regression line between reference values and estimates of Eq. (1) did not pass through the origin. An ordinary linear regression could make the regression line closer to the origin, but a Passing & Bablok regression could just make it pass through the origin (Fig. 4). The Bland–Altman plots confirm the advantages of using the Passing & Bablok regression in this study. In Fig. 5, compared to the Passing & Bablok regression, an ordinary linear regression can make the bias between estimated and reference values closer to 0, but can obviously enlarge the agreement range. Among all recruits, before and after the Passing & Bablok regression, the difference between estimates and reference values for the left kidney decreased from 0.17±0.70 (t=3.13, P=0.00) to 0.01±0.56 cm (t=0.19, P=0.848) and that for the right kidney decreased from 0.17±0.77 (t=2.82, P=0.005) to 0.004±0.64 cm (t=0.09, P=0.93), respectively. Therefore, the test formula was Eq. (1) corrected by the Passing & Bablok regression (Eqs. (2) and (3)).

Left kidney depth = 1.12022 + 14.8162 × W/H – 0.01353 × A

Right kidney depth = 1.07943 + 16.0939 × W/H – 0.02476 × A

3.2. Validation and comparison of renal depth formulas

The test formula regressed from group A, together with the 4 “standard” formulas, was evaluated in the 2 groups by the indices described in the statistical analysis section. All indices of the test formulas were better than other formulas (Fig. 6A–J), and P5 mm for left and right kidneys were 63.0% and 59.3% in group A and 66.0% and 71.7% in group B.

However, between the 2 groups, the indices of the 4 “standard” formulas were not concordantly indicated by the best formula. In summary, a higher accuracy did not always mean a better precision, and results of P5 mm were different from other accuracy indices (bias and absolute bias). Because of the practical meaning of P5 mm, it was considered as the reasonable index in this study, and other indices acted to assist in getting information of systematic bias and random error (refer to the “Discussion” section).

Compared to other “standard” formulas, P5 mm of the Gordon formula was the highest in predicting right kidney depth in group A (49.4%) and group B (49.1%). However, P5 mm of the Raynaud (52.5%) and the Taylor formulas (60.4%) were the highest in predicting left kidney depth in groups A and B, respectively. Although P5 mm of the Gordon formula was superior under some conditions, its precision was not so good enough (correlation coefficient and RMSE were 0.897 and 0.904 cm for left kidney of group A, respectively). The accuracy of Raynaud formula, which was also regressed from children, only was better than the Gordon formula in estimating left kidney depth of group A (Fig. 6E–J).

Above all, compared to those of the “standard” formulas, all indices of the test formula were obviously improved in the 2 groups.
3.3. Discriminatory ability of the test formula

In predicting left and right kidney depths of all recruits, 78 (36.3%) and 81 (37.7%) estimates of the test formula (Eqs. (2) and (3)) were outside the P5 mm range, respectively. Among age, body weight, height, group, measured kidney depth, and BMI, no variable was significant for left kidney outliers, and only age was significant for right kidney outliers. OR and Wald of the right kidney outliers model were 1.162 (95% CI: from 1.040 to 1.162) and 16.386 (P = 0.000), respectively. However, age alone had little influence in the discrimination, and the c statistic was 0.634 (95% nonparametric CI: 0.559–0.709).

We further analyzed the influence of age on our kidney depth formula. As shown in Fig. 1B, the P5 mm less than 50% frequently corresponded with smaller sample size. Among the recruits whose age was less than 10 years (≤ 10 years), P5 mm for left and right kidney depth formulas were 70.1% (47/67) and 76.1% (51/67), respectively. Among the recruits aged from 11 to 14 years, P5 mm of left and right kidney depth formula were 55.6% (25/45) and 68.9% (31/45), respectively. Among the recruits older than 15 years (≥ 15 years), P5 mm of left and right kidney depth formula were 63.1% (65/103) and 50.5% (52/103), respectively.

4. Discussion

Our results indicated that in predicting pediatric kidney depth, accuracy and precision of a step-by-step regressed formula were
better than the 4 “standard” equations. The probability of its estimates within a range of -3 to 5 mm was about 60% among children aged from 1 to 19 years, and the probability was even more than 70% among children younger than 10 years.

In the fields of science and statistics, accuracy and precision are not replaceable concepts. Accuracy is associated with systematic error (inherent limitation of a formula), and precision is associated with random error (inconsistent when repeated). It is most important to note that a higher accuracy does not always mean a higher precision. As illustrated by our data, the “standard” formulas could not concordantly decrease the 2 types of errors and did not have both higher accuracy and precision. However, all indices of the test formula were improved in groups A and B and strongly suggested that the step-by-step regressed formula was better than others in predicting pediatric kidney depth.

In previous studies, bias and absolute bias are frequently used accuracy indices, but the 2 indices cannot give us the information on the bias range of estimates. Therefore, to better meet our interest in the estimation, P5 mm is superior to other accuracy indices for defining a bias range. In other words, it can declare the question that how frequently the differences between measured values and estimates are in the range.

Although P5 mm could indicate the best formula in this study, as an accuracy index, it was necessary to evaluate the precision of formulas. According to our results, P5 mm of the Taylor formula for left kidney depth was 60.4% in group B, but the value decreased to 43.8% in group A (Fig. 6I and J). Therefore,

![Figure 3](image1.png)

**Figure 3.** Estimated coefficients and 95% CI of Eq. (1) in the groups A and B. Panels (A) and (B) are for the coefficient of a and b, respectively.

![Figure 4](image2.png)

**Figure 4.** Scatter plots of measured values against estimates of test formulas in the group A. The left and right columns are for left kidney (A, C, and E) and right kidney (B, D, and F), respectively. From the top to the bottom, rows are plots for the values directly estimated by Eq. (1) (A and B), Eq. (1) with ordinary linear regression (C and D), and Eq. (1) with Passing & Bablok regression (E and F), respectively.
accuracy of the Taylor formula was inconsistent when repeated. Our results also supported the recommendation of Xeleris workstation that both the Gordon and the Raynaud formulas could be used for the estimation. Although P5 mm of the Gordon formula was better in some populations, its precision was not so good enough ($r = 0.879$ to $0.913$ in groups A and B). Its lower precision could explain why P5 mm of the Gordon formula was slightly worse than the Raynaud formula in other populations. Above all, to avoid the false conclusion resulting from random error, it is necessary to validate the formula in different populations.

Defining errors to be in a range of $-5$ to $5$ mm appears to be strict, but it can meet our requirement for camera-based GFR measurement. The P5 mm range did not take the measurement deviation into account as done for P30 (the proportion of estimates within 30% deviation of measurement). Further in predicting left and right kidney depths, P30 of the test formula was both as high as 99.1% (213/215), and P15 of the formulas were 82.3% (177/215) and 84.2% (181/215), respectively. Therefore, it is possible that P15 or P30 might not identify the performance of various formulas with slight differences, and P5 mm is more suitable for the evaluation.

Although the general form of our pediatric formula (Eqs. (2) and (3)) was the same as the Taylor, they were regressed by different methods. The Taylor formula was fitted by multiple linear stepwise regression, and ours was regressed through nonlinear regression and subsequently corrected by the Passing & Bablok regression. Because the coefficients of our formula were narrower and overlapped 95% CIs, its repeatability was much better among different populations. The Passing & Bablok regression was used, because if the estimates were closer to their reference values, their regression line would pass through the origin with a slope of 1. Before and after the PBR correction, both mean bias and standard difference decreased among all recruits, and the statistical significance between estimates and reference values reduced. Therefore, as indicated by our data, the Passing & Bablok regression is better than the ordinary linear regression in fitting kidney depth formulas. Besides the same general form, most coefficient values between the Taylor formula and ours only had slight difference. Because the 2 formulas were separately regressed from children and adults, the slight differences indicated that the general form might be applied in all individuals regardless their age and should be verified in future. However, there were some differences between the Taylor formula and ours. For example, the coefficients of age were negative in our formula for both left and right kidneys, but only the constant of Taylor right kidney depth formula was negative. The difference could explain the different accuracy and precision between the Taylor formula and ours.

In the validation, the coefficients of formulas from the groups A and B were in narrower and overlapped 95% CIs and indicated that systematic differences of the 2 test formulas were not significantly different. Therefore, in our opinion, the test formula could be applied in other pediatric populations with higher accuracy and higher precision. The test formula was regressed from the recruits younger than 19 years and was obviously better than the “standard” formulas. However, further analysis indicated that the formula did not perform equally well throughout the age range. Both the left and the right kidney depth formulas were highly reliable among the children younger than 10 years. The performance of left kidney depth formula was not so good among children aged from 11 to
14 years and might be explained by the limited sample size which resulted in the sampling error. However, for the right kidney depth formula, the performance decreased obviously among children older than 15 years. In summary, the test formula performed well among children, especially those younger than 10 years. The performance of the test formula for the older children should be verified in future.

In addition, according to a study from Larson et al.,[18] in measuring renal length, the absolute errors of CT measurement were only $0.72 \pm 0.78$ mm, and the interobserver errors were $0.88 \pm 0.81$ mm. Therefore, for developing formulas to estimate kidney depth, CT measurement could be considered as the reference method.[19]

Above all, the step-by-step regressed formula performed the best in our recruits with a higher repeatability, accuracy, and precision. Furthermore, the general form of our formula has the possibility to be used for both children and adults in estimating kidney depth.

Figure 6. Validation of formulas. The test formulas regressed from the group A are validated in the 2 groups. The indices of mean bias (A and B), absolute bias (C and D), correlation coefficient (E and F), root-mean square error (G and H), and P5 mm (I and J) are plotted for the 2 groups. The left and right columns are the plots for the groups A and B, respectively.
5. Conclusion
To predict pediatric kidney depth, accuracy and precision of a step-by-step regressed formula were better than others, especially among children younger than 10 years.

References
[1] Gordon I, Piepsz A, Sixt R. Guidelines for standard and diuretic renogram in children. Eur J Nucl Med Mol Imaging 2011;38:1175–88.
[2] Taylor AT. Radionuclides in nephrourology, Part 2: Pitfalls and diagnostic applications. J Nucl Med 2014;55:786–98.
[3] Lythgoe MF, Gradwell MJ, Evans K, et al. Estimation and relevance of depth correction in paediatric renal studies. Eur J Nucl Med 1998;25:115–9.
[4] Gruenewald SM, Collins LT, Fawdry RM. Kidney depth measurement and its influence on quantitation of function from gamma camera renography. Clin Nucl Med 1985;10:398–401.
[5] Pasquarriello A, Innocenti M, Ratti V, et al. Theoretical calculation of optimal depth in the percutaneous native kidney biopsy to drastically reduce bleeding complications and sample inadequacy for histopathological diagnosis. Nephrol Dial Transplant 2007;22:5516–20.
[6] Granerus G, Moonen M. Effects of extra-renal background subtraction and kidney depth correction in the measurement of GFR by gamma camera renography. Nucl Med Commun 1991;12:519–27.
[7] Gordon I, Evans K, Peters AM, et al. The quantitation of 99Tcm-DMSA in paediatrics. Nucl Med Commun 1987;8:661–70.
[8] Raynaud C, Jacquot C, Freeman LM. Measuring renal uptake of 197HGCl2 by gamma camera. Radiology 1974;110:413–8.
[9] Tomnesen KH, Munch O, Hald T, Winkel K, Blaufox MD, Funck-Brentano JL. Influence on the renogram of variation skin to kidney distance and the clinical importance hereof. Proceedings of the Third International Symposium on Radionuclides in Nephrology Thieme, Stuttgart:1974;79–86.
[10] Taylor A, Lewis C, Giacometti A, et al. Improved formulas for the estimation of renal depth in adults. J Nucl Med 1993;34:1766–9.
[11] Ma G, Shao M, Xu B, et al. Establish new formulas for the calculation of renal depth in both children and adults. Clin Nucl Med 2015;40:e357–62.
[12] De Bievre P. The 2007 international vocabulary of metrology (VIM), JCGM 200: 2008 [ISO/IEC Guide 99]: meeting the need for intercontinentally understood concepts and their associated intercontinentally agreed terms. Clin Biochem 2009;42:246–8.
[13] Schaeffter ES, Ebert N, Delanaye P, et al. Two novel equations to estimate kidney function in persons aged 70 years or older. Ann Intern Med 2012;157:471–81.
[14] Passing H, Bablok . A new biometrical procedure for testing the equality of measurements from two different analytical methods. Application of linear regression procedures for method comparison studies in clinical chemistry, Part I. J Clin Chem Clin Biochem 1983;21:709–20.
[15] Si H, Lei Z, Li S, et al. Lean body mass is better than body surface area in correcting GFR. Clin Nucl Med 2013;38:e210–5.
[16] Carlier M, Dumoulin A, Janssen A, et al. Comparison of different equations to assess glomerular filtration in critically ill patients. Intens Care Med 2015;41:427–35.
[17] Venkatesh B, Ruddington DW. Measuring PCO2 for gastric tonometry: precision is as important as bias. Anaesthesia 1995;50:913.
[18] Larson DB, Meyers ML, O’Hara SM. Reliability of renal length measurements made with ultrasound compared with measurements from helical CT multiplanar reformat images. Am J Roentgenol 2011;196:W592–7.
[19] Prigent A, Cosgriff P, Gates GF, et al. Consensus report on quality control of quantitative measurements of renal function obtained from the renogram: International Consensus Committee from the Scientific Committee of Radionuclides in Nephrourology. Semin Nucl Med 1999;29:146–59.