A comparison of cystatin C- and creatinine-based prediction equations for the estimation of glomerular filtration rate in black South Africans

Hendrick E. van Deventer¹, Janice E. Paiker¹, Ivor J. Katz² and Jaya A. George¹

¹Department of Chemical Pathology and NHLS, University of the Witwatersrand, Johannesburg, South Africa and ²Division of Nephrology, Chris Hani Baragwanath Hospital, University of the Witwatersrand, Johannesburg, South Africa

Abstract

Background. Serum creatinine (S-Cr)-based prediction equations are commonly used for estimating glomerular filtration rate (GFR). However, S-Cr concentration is also affected by other factors such as tubular secretion, muscle mass, diet, gender and age. Serum cystatin C (S-Cys C)-based prediction equations have been proposed as an improved potential alternative as S-Cys C levels are not influenced by many of the factors that affect creatinine concentration other than GFR. This may be of great benefit to patients with low muscle mass such as those infected with human immunodeficiency virus who are at increased risk for the development of renal impairment. The aim of this study was to develop and evaluate a S-Cys C-based prediction equation for different stages of renal disease in black South Africans.

Methods. One hundred patients with varying degrees of renal function were enrolled in the study. The plasma clearance of ⁵¹Cr-EDTA, a gold standard method, was used to measure GFR (mGFR). In addition, serum was analysed for S-Cr and S-Cys C on each participant. This dataset was split into a development dataset (n = 50) and a test dataset (n = 50). The development dataset was used to formulate a S-Cys C- and S-Cr-based prediction equation using multiple linear regression analysis. These equations together with the four-variable MDRD and CKD-EPI equation were then tested on the test dataset.

Results. In the test dataset, accuracy within 15% of measured GFR was 68% for the S-Cys C equation and 48% for the S-Cr equation. Root mean square error for S-Cys C eGFR was 10.2 mL/min/1.73 m² for those patients with mGFR <60 mL/min/1.73 m² and 11.9 mL/min/1.73 m² for those patients with mGFR >60 mL/min/1.73 m². Root mean square error for S-Cr eGFR was 10.7 mL/min/1.73 m² for those patients with mGFR <60 mL/min/1.73 m² and 25.5 mL/min/1.73 m² for those patients with mGFR >60 mL/min/1.73 m².

Conclusions. In this study, S-Cys C-based prediction equations appear to be more precise than those of S-Cr for those patients with mGFR >60 mL/min/1.73 m² and may therefore be of benefit in the earlier detection of renal impairment.

Keywords: creatinine; cystatin C; glomerular filtration rate; MDRD

Introduction

Glomerular filtration rate (GFR) is considered the ‘gold standard’ in the diagnosis of chronic kidney disease (CKD) and is also accepted as the best overall measure of kidney function [1,2]. GFR can be measured as the renal clearance of exogenous markers such as inulin, ⁵¹chromium ethylenediaminetetraacetic acid (⁵¹Cr-EDTA), technetium-labelled diethylene-triamine-pentacetate (⁹⁹ᵐTc-DTPA) and iohexol. However, these exogenous markers are impractical for routine clinical use due to their limited access and high cost. Endogenous GFR markers include serum creatinine (S-Cr) and cystatin C (S-Cys C). S-Cr is the most commonly used marker in the clinical laboratory to assess GFR; however, it has multiple limitations [3]. For example, creatinine concentration is
not only determined by GFR but also affected by factors such as muscle mass, diet, gender and age [4,5]. This results in a large intra-individual variation in creatinine production. The four-variable Modification of Diet in Renal Disease (4-v MDRD) and Cockcroft–Gault (CG) equations, two S-Cr-based equations commonly used for estimating GFR, account for some of these factors. The Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation, also based on S-Cr, has recently been proposed as a more accurate and precise estimation of GFR especially at higher GFR levels [6]. However, the intra-individual variation in creatinine production still leads to imprecision in these GFR estimation equations [7–9]. These equations are not also considered accurate for detecting early kidney disease [10,11], and a few studies evaluated equations for estimating GFR in patients with GFR >60 mL/min/1.73 m².

S-Cys C possesses many of the attributes required of the ideal GFR marker. It is a low-molecular-weight protein (13.3 kDa) produced by all nucleated cells [12], and is almost completely filtered by the renal glomerulus and normally almost completely reabsorbed and degraded by proximal tubular cells. S-Cys C therefore correlates closely with GFR [13]. Many recent studies have compared S-Cr-based prediction equations with S-Cys C-based prediction equations and have found them comparable [14,15]. Importantly, S-Cys C levels do not show a strong dependence as S-Cr on muscle mass, age and gender [16]. The relative independence on muscle mass and diet may be of great benefit to black South Africans as many are from a lower socioeconomic group and diseases such as human immunodeficiency virus (HIV) have a high prevalence. However, S-Cys C-based prediction equations for the detection of renal disease have not previously been evaluated in this population.

The aim of this study was 3-fold, namely (i) to develop S-Cys C- and S-Cr-based prediction equations for estimating GFR using 51Cr-EDTA plasma clearance as the reference mGFR, (ii) to compare the developed S-Cys C and S-Cr prediction equations to measured GFR in a cohort of black South Africans, and (iii) to determine whether S-Cys C prediction equations offer any advantage over S-Cr-based prediction equations in this population.

Materials and methods

Participants

During 2006, one hundred participants were recruited to take part in the study after being screened and counselled by their clinicians. All participants gave written informed consent after being educated with regard to the potential benefits, risks and the study procedures. The study was approved by the Human Research Ethics Committee of the University of the Witwatersrand (Protocol number: M060137).

Participants were either inpatients at the Chris Hani Baragwanath hospital or being followed up at the renal unit outpatient department at the hospital. Participants were identified by the treating physician as being at risk for the development of chronic kidney disease as well as patients with established CKD. All participants were black South Africans older than 18 years. Exclusion criteria were pregnancy, acute kidney injury and oedema. A total of 51 males and 49 females were enrolled in the study. The participants suffered from different diseases, the most common of which were hypertension (n = 36), diabetes mellitus (n = 25) and HIV (n = 20). Participants being worked up for possible kidney donation were also included (n = 7). GFR, S-Cr and S-Cys C were measured for all 100 participants.

Test methods

Age (years), standing height (centimetres), weight (kilogram) and gender were recorded for all participants. A 5-mL blood sample was collected for S-Cr and S-Cys C measurement on the same day when GFR measurement was performed. GFR was measured as the plasma clearance of 51Cr-EDTA which has been shown to be a reliable and simple method for measuring GFR [17–19]. GFR was measured with the slope intercept method [17] corrected with the Brochner–Mortensen equation [20] according to guidelines adopted by the British Nuclear Medicine Society [21]. GFR was normalized to body surface area (BSA) using the DuBois method (BSA (m²) = [71.84 × weight (kg)⁰.⁴²⁵ × height (cm)⁰.⁷²⁵] / 10 000) [22]. In this paper, this will be known as measured GFR (mGFR).

S-Cr was measured using an alkaline picrate assay (Roche Modular analyser, Roche Diagnostics, Mannheim, Germany) traceable to isotope dilution mass spectrometry (IDMS). Cystatin C was measured by an automated particle-enhanced immunoturbidimetric (PETIA) assay (DAKO Cytomation) on a Roche Modular analyser system with reagents according to the recommended guidelines by the manufacturer. To assess possible calibration differences, a calibration panel obtained from the Cleveland Clinic Foundation was used. This calibration panel consisted of 40 serum samples, with values assigned by a Roche enzymatic assay (Creatinine Plus, Roche Diagnostics) measured at the Cleveland Clinic Laboratories. This assay has been independently validated to be traceable to IDMS [23].

Model development and evaluation

A random number generator (Random.org) was used to select 50 of the 100 results to create a development dataset. Least-square linear regression analysis was applied to this dataset to develop a S-Cys C equation (S-Cys C eGFR) to estimate GFR. In order to compare S-Cys C and S-Cr eGFR equations, least-square linear regression analysis was also used to develop a S-Cr-based equation (S-Cr eGFR). Variables evaluated for inclusion in these equations were age, gender, height and weight. Variables were included in the prediction equations only if their P-value was <0.05. GFR was log-transformed to equalize the variance across the range of GFR. Both these equations and the previously established 4-v MDRD equation without the ethnicity factor for race [24] and the CKD-EPI equation [6] both with and without the ethnicity factor for race were then evaluated in the remaining 50 participants (test dataset).

Statistical analysis

Statistical analysis was conducted using Analyse-it for Microsoft Excel. The Shapiro–Wilk test was used to assess normality. Non-parametric continuous data variables are expressed as median (interquartile range). The performance of the S-Cys C eGFR, S-Cr eGFR, 4-v MDRD and CKD-EPI prediction equations relative to mGFR were assessed in the test dataset using Pearson correlation coefficient ($R^2$), Passing and Bablok regression analysis, median bias, interquartile range (IQR), root mean square error (RMSE), and accuracy within 15% and 30% ($P_{15}$ and $P_{30}$) of mGFR. Results were then ranked according to mGFR, and the dataset were split into two groups: GFR <60 mL/min/1.73 m² and GFR ≥60 mL/min/1.73 m². Bias, IQR and RMSE were then calculated in each of these groups.

Results

Creatinine comparison to enzymatic assay

Evaluation of S-Cr calibration was based on 39 observations as one of the samples with a difference between the assigned value and the measured value of more than 3 SDs from the mean difference was excluded from the analysis. The measurements were done in triplicate in three separate runs with measured S-Cr values ranging from 44 to 398 μmol/L. The correlation between the Cleveland Clinic Foundation (CCF)-assigned values and the South African (SA) measured values was high ($R^2 = 0.999$). Deming regression analysis was used to calculate
the slope [0.964 (95% CI 0.952–0.975)] and intercept [0.039 (95% CI 0.010–0.068)] of the regression equation, with \( y = \text{CCF-assigned values} \) and \( x = \text{SA measured values} \).

*eGFR prediction equations*

A S-Cys C-based GFR prediction equation including the variables S-Cys C and age (S-Cys C eGFR) was developed (Table 1). The variables, gender, weight and height did not significantly affect the equation (P = 0.73, P = 0.19 and P = 0.40, respectively) and were therefore excluded from the calculation. Logarithmic transformation of GFR resulted in normal distribution of residuals about the zero line. Multiple linear regression analysis on logarithmic transformation of S-Cys C and age did not yield better results. The S-Cys C-based prediction equation developed from this dataset is therefore: S-Cys C eGFR = \( 10^{2.35 \times 10^{-0.33 \times \text{S-Cys C (mg/L)}} \times -0.003 \times \text{Age}} \).

<table>
<thead>
<tr>
<th>Prediction equation</th>
<th>Intercept</th>
<th>Log S-Cr (μmol/L)</th>
<th>If female</th>
<th>S-Cys C (mg/L)</th>
<th>Age (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-Cys C eGFR</td>
<td>2.35 (0.05)</td>
<td>-1.12 (0.06)</td>
<td>-0.33 (0.01)</td>
<td>-0.0029 (0.0010)</td>
<td></td>
</tr>
<tr>
<td>S-Cr eGFR</td>
<td>4.21 (0.11)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Coefficients for predicting log (GFR) derived from multiple linear regression analysis in this dataset of 50 black South Africans. SE, standard error.

Fig. 1. Bland–Altman percentage difference plots. (A) S-Cr eGFR. (B) S-Cys C eGFR.
The limits of agreement were evaluated without including this factor. The 95% factor for race was included, the CKD-EPI equation served in the CKD-EPI equation when the correction factor established for African Americans was included in the equation \([19.5\% (12.4–26.6\%)]\). However, without the ethnicity factor, the bias for the CKD-EPI equation improved to 5.0% \((-5.19–5.54\%\) for predicting \(\text{mGFR} < 60 \text{ mL/min/1.73 m}^2\) (P = 0.22). The AUCs for the 4-v MDRD \(0.94 (0.88–1.00)\) and CKD-EPI \(0.99 (0.97–1.00)\) were significantly better than for S-Cys alone \(0.87 (0.76–0.98)\) for predicting \(\text{mGFR} < 60 \text{ mL/min/1.73 m}^2\) (P < 0.05) (Table 2).

For those patients with \(\text{mGFR} > 60 \text{ mL/min/1.73 m}^2\), the IQR of the difference between estimated and measured GFR was 53.7–55.8% for the S-Cys equation and 37.5–39.3% for the S-Cys C eGFR equation (Figure 1). Accuracy within 30% was 74% for the 4-v MDRD eGFR equation, 72% for the CKD-EPI equation, 70% for the S-Cr equation and 84% for the S-Cys C eGFR.

Receiver operating characteristic (ROC) curves revealed a similar area-under-the-curve (AUC) for all four equations for predicting \(\text{mGFR} < 60 \text{ mL/min/1.73 m}^2\). There was no significant difference between the AUC for Cys C on its own \([0.97 (0.93–1.00)]\) compared with the S-Cys C equation \([0.99 (0.97–1.00)]\) for predicting \(\text{mGFR} < 60 \text{ mL/min/1.73 m}^2\) \((P = 0.22)\). The AUCs for the 4-v MDRD \(0.94 (0.88–1.00)\) and CKD-EPI \(0.96 (0.90–1.00)\) were significantly better than for S-Cr alone \(0.87 (0.76–0.98)\) for predicting \(\text{mGFR} < 60 \text{ mL/min/1.73 m}^2\) (P < 0.05) (Table 2).

**Evaluation of eGFR prediction equations**

Bland–Altman percentage difference plots showed a significant bias for the CKD-EPI equation when the ethnicity factor established for African Americans was included in the equation \([19.5\% (12.4–26.6\%)]\). However, without the ethnicity factor, the bias for the CKD-EPI equation improved to 5.0% \((-2.2–12.2\%)\). Because of the bias observed in the CKD-EPI equation when the correction factor for race was included, the CKD-EPI equation was evaluated without including this factor. The 95% limits of agreement were \(-49.7–52.3\%\) for the 4-v MDRD eGFR equation, \(-45.3–62.4\%\) for the CKD-EPI equation, \(-53.7–55.8\%\) for the S-Cys C eGFR equation and \(-37.5–39.3\%\) for the S-Cys C eGFR equation (Table 1). Accuracy within 30% was 74% for the 4-v MDRD eGFR equation, 72% for the CKD-EPI equation, 70% for the S-Cr equation and 84% for the S-Cys C eGFR.

Multiple linear regression analysis was also used to develop a S-Cr-based prediction equation from this dataset (S-Cr eGFR) (Table 1). S-Cr eGFR = 10\(^{4.21}\) × (S-Cr (μmol/L))\(^{-1.12}\) × 0.78 if female.

### Table 2. Performance of equations in different stages of renal disease: bias and precision

<table>
<thead>
<tr>
<th>eGFR</th>
<th>Median bias (95% CI) (mL/min/1.73 m(^2))</th>
<th>IQR(^a) (mL/min/1.73 m(^2))</th>
<th>RMSE(^b) (mL/min/1.73 m(^2))</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>4-v MDRD eGFR</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\text{mGFR} (&lt; 60 \text{ mL/min/1.73 m}^2))</td>
<td>(-3.1) ((-7.0–3.4))</td>
<td>10.6</td>
<td>11.0</td>
</tr>
<tr>
<td>(\text{mGFR} (&gt; 60 \text{ mL/min/1.73 m}^2))</td>
<td>(0.6) ((-8.0–12.5))</td>
<td>28.2</td>
<td>28.2</td>
</tr>
<tr>
<td>Overall</td>
<td>(-1.5) ((-5.4–2.5))</td>
<td>20.5</td>
<td>20.5</td>
</tr>
<tr>
<td><strong>CKD-EPI GFR</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\text{mGFR} (&lt; 60 \text{ mL/min/1.73 m}^2))</td>
<td>(-2.3) ((-5.5–7.2))</td>
<td>12.7</td>
<td>12.9</td>
</tr>
<tr>
<td>(\text{mGFR} (&gt; 60 \text{ mL/min/1.73 m}^2))</td>
<td>(11.0) ((-2.9–21.4))</td>
<td>24.3</td>
<td>24.6</td>
</tr>
<tr>
<td>Overall</td>
<td>(4.9) ((-2.8–12.3))</td>
<td>22.4</td>
<td>23.0</td>
</tr>
<tr>
<td><strong>S-Cr eGFR</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\text{mGFR} (&lt; 60 \text{ mL/min/1.73 m}^2))</td>
<td>(-2.0) ((-4.8–5.4))</td>
<td>10.5</td>
<td>10.7</td>
</tr>
<tr>
<td>(\text{mGFR} (&gt; 60 \text{ mL/min/1.73 m}^2))</td>
<td>(-3.8) ((-10.7–3.9))</td>
<td>25.2</td>
<td>25.5</td>
</tr>
<tr>
<td>Overall</td>
<td>(-2.0) ((-6.0–1.6))</td>
<td>14.9</td>
<td>15.1</td>
</tr>
<tr>
<td><strong>S-Cys C eGFR</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\text{mGFR} (&lt; 60 \text{ mL/min/1.73 m}^2))</td>
<td>(-1.8) ((-4.9–4.9))</td>
<td>10.1</td>
<td>10.2</td>
</tr>
<tr>
<td>(\text{mGFR} (&gt; 60 \text{ mL/min/1.73 m}^2))</td>
<td>(1.3) ((-4.0–4.8))</td>
<td>11.8</td>
<td>11.9</td>
</tr>
<tr>
<td>Overall</td>
<td>(-0.3) ((-2.6–3.4))</td>
<td>10.6</td>
<td>10.6</td>
</tr>
</tbody>
</table>

\(^a\)Interquartile range of the difference between estimated and measured GFR.

\(^b\)Root mean square error calculated as the square root of the \([\text{median difference in estimate} – \text{measured}]^2 + (\text{interquartile range of the difference})^2\).
was 11.8 mL/min/1.73 m² for the S-Cys C eGFR and 25.2 mL/min/1.73 m² for S-Cr eGFR. For those patient with mGFR <60 mL/min/1.73 m², the IQR of the difference between estimated and measured GFR was 10.1 mL/min/1.73 m² for the S-Cys C eGFR and 10.5 mL/min/1.73 m² for S-Cr eGFR (Table 3 and Figure 2).

Evaluation of eGFR prediction equations in patients known to be HIV positive

The study population also included 20 participants known to be infected with HIV; of these, 11 participants were included in the development dataset, and 9 participants in the test dataset. The performances of all four equations were evaluated in all 20 participants. The CKD-EPI equation significantly overestimated GFR in this group of patients. Pearson correlation coefficient was 0.81 for S-Cys C eGFR and 0.77 for the 4-v MDRD and S-Cr eGFR equations. Accuracy within 30% was 65% for the 4-v MDRD equation, 65% for the S-Cys C eGFR equation and 70% for the S-Cr eGFR equation. IQR was 16.1 mL/min/1.73 m² for the 4-v MDRD equation, 17.7 mL/min/1.73 m² for the S-Cr eGFR equation and 14.9 mL/min/1.73 m² for the S-Cys C eGFR equation (Supplemental Table 1).

Discussion

In this study, S-Cr and S-Cys C equations demonstrated a wide scatter of eGFR around the ‘gold standard’ measured GFR.

Imprecision between measured and estimated GFR, as assessed by calculating the IQR between measured and estimated GFR, was smaller for the S-Cys C-based eGFR prediction equation than S-Cr-based eGFR prediction equation for those patients with mGFR >60 mL/min. A potential role of S-Cys C eGFR prediction therefore seems to be the more precise estimation of GFR for those patients with early kidney disease. S-Cys C-based prediction equations may play a potential role in the detection of early kidney disease in those patients at increased risk for the development of CKD, such as patients with diabetes mellitus (DM), hypertension (HT) and HIV.

Factors such as gender, weight and height did not have a statistically significant impact on this study and were therefore excluded from the S-Cys C prediction equation. Age, although statistically significant for inclusion, contributed minimally to the S-Cys C eGFR equation. A potential reason for this may be the relative independence of muscle mass on S-Cys C concentration [12]. In HIV-infected patients with decreased body mass, creatinine may underestimate the prevalence of CKD [25]. The relative independence of S-Cys C on muscle mass may therefore be a potential advantage in patients with decreased muscle mass such as those infected with HIV. This study included 20 participants known to be infected with HIV, and 15 participants had a BMI <20 kg/m². In the group of participants known to be infected with HIV, accuracy within 30% of mGFR was best for the S-Cys C eGFR equation, and IQR was 14.9 mL/min/1.73 m² for the S-Cys C equation compared with 17.7 mL/min/1.73 m² for the S-Cr eGFR equation. These data suggest that S-Cys C eGFR equations may be of potential benefit in this group of patients. Further studies are however needed.

Important limitations to the use of S-Cys C immunoassay determinations lie in the lack of standardization of results and the high cost of performing the assay [26]. Currently, S-Cys C is not listed on the Joint Committee for Traceability in Laboratory Medicine (JCTLM) database [27]. Because of differences in results obtained with different methods, no universal equation or cut-off value can be used. Furthermore, most of the assay kits currently available for S-Cys C measurement are performed as a batch analysis, which significantly delays the turnaround time of its measurement. This is in contrast to that of
S-Cr measurement, which is mostly automated, allowing for quick and easy determination of the renal status of patients.

An important limitation to this study is the small dataset used for the development and testing of both S-Cr and S-Cys C equations.

In conclusion, in this study, a S-Cys C-based prediction equation was developed and tested using $^{51}$Cr-EDTA as the reference mGFR in a cohort of black South African patients. This cohort included patients with various causes of CKD, namely HT, DM and HIV. Imprecision between measured and estimated GFR was smaller for the S-Cys C equation estimated also more patients within 30% of mGFR than any of the other equations. A benefit of S-Cys C-based prediction equations over that of the creatinine-based equations may be the more precise estimation of GFR for patients with early kidney disease in black South Africans.

**Supplementary data**

Supplementary data is available online at http://ndt.oxfordjournals.org.

**Acknowledgements.** We thank the Department of Nuclear Medicine for performing the $^{51}$Cr-EDTA measurements.

**Conflict of interest statement.** None declared.

**References**

22. Dubois D, Dubois EF. A formula to estimate the approximate surface area if height and weight be known. *Arch Intern Med* 1916; 17: 863–871