Error Estimates for Calculated Glomerular Filtration Rates

Simon Brown

Abstract—Glomerular filtration rate (GFR) is a measure of kidney function. It is usually estimated from serum concentrations of cystatin C or creatinine although there has been considerable debate in the literature about (i) the best equation to use and (ii) the variability in the correlation between the concentrations of creatinine and cystatin C. The equations for GFR can be written in a general form and from these I calculate the error of the GFR estimates associated with analyte measurement error. These show that the error of the GFR estimates is such that it is not possible to distinguish between the equations over much of the concentration range of either analyte. The general forms of the equations are also used to derive an expression for the concentration of cystatin C as a function of the concentration of creatinine. This equation shows that these analyte concentrations are not linearly related. Clinical reports of cystatin C and creatinine concentration are consistent with the expression derived.

Keywords—creatinine, cystatin C, error analysis, glomerular filtration rate, measurement error.

I. INTRODUCTION

Glomerular filtration rate (GFR) is a well established measure of kidney function. Values decline with age and depend on gender, but are about 120-130 mL min\(^{-1}\) (1.73 m\(^2\))\(^{-1}\) in healthy young adults. In chronic kidney disease (CKD) GFR declines and reaches values of less than 15 mL min\(^{-1}\) (1.73 m\(^2\))\(^{-1}\) in stage 5 CKD. Estimates of GFR are often used in the detection, evaluation and management of CKD [7].

The GFR is the volume of fluid filtered through the glomerular capillaries of the kidney into Bowman’s capsule in a given time. This can be measured by comparing the concentration of an analyte in the blood (C\(_\text{blood}\)) and the urine (C\(_\text{urine}\)) for a given urine volume (V\(_\text{urine}\), which has units of mL min\(^{-1}\))

\[
\text{GFR} = \frac{C_{\text{urine}} V_{\text{urine}}}{C_{\text{blood}}}
\]  

(1)

and it is usually expressed per unit body surface area (so the usual units are mL min\(^{-1}\) (1.73 m\(^2\))\(^{-1}\)), which can be estimated in m\(^2\) from height (h, in cm) and weight (w, in kg) using

\[
A_{\text{body}} = 0.007184 h^{0.725} w^{0.425}
\]

(2)

[10], for example. Several analytes have been used to estimate GFR, but the distribution of exogenous compounds such as inulin, diethylenetriamine pentaacetic acid or iothalamate [14] are expensive and difficult to employ in normal clinical practice [7]. Instead, endogenous markers, including creatinine and cystatin C, have been used. Unfortunately, the concentration of these analytes varies with age, gender and physiological conditions. Creatinine, especially, has complex chemistry and biochemistry [15, 16], the concentration varies with muscle mass and physical activity [17] and assay results can be affected by the presence in the sample of common drugs [18]. Moreover, it has been argued that creatinine is not a sensitive indicator of GFR [19-21]. Cystatin C concentration is also variable [22], but the variation with age and gender is less pronounced than that reported for creatinine in at least some cases [23]. The correlation between the concentrations of creatinine and cystatin C can be weak (r = 0.18) [22-25] or relatively strong (r = 0.9) [26].

Many equations are used to estimate GFR, but they are all based on laboratory measurements of either cystatin C or serum creatinine, although the latter expressions often involve other measurements (usually age and weight, but also serum urea nitrogen or albumin). Several recent publications have compared these equations and, in most cases, one equation is identified as superior to the others [3, 27-32].

As one would anticipate, all the comparable equations yield similar estimates. This prompts one to ask whether they can be distinguished given the uncertainty of the laboratory measurements on which they are based. The equations are based on statistical analysis, but it appears that each input analyte value is treated as error-free.

No analytical estimate of the uncertainty of the GFR estimates has been reported in the literature. This omission is rectified here. I show that the uncertainty of the GFR estimates arising just from the laboratory measurements makes them effectively indistinguishable. It is, therefore, not reasonable to suggest that any one equation is more reliable than any other until more precise analytical techniques become available. I also show that the variability in the correlation between cystatin C and creatinine concentration is consistent with the GFR equations.

II. GENERAL FORM OF THE EQUATIONS

To simplify the analysis, and for clarity, I write the GFR equations in general form. This makes it possible to analyse the various equations based on a specific analyte once, which
TABLE I

COEFFICIENT VALUES FOR SOME CYSTATIN C-BASED ESTIMATES OF GFR (3)

<table>
<thead>
<tr>
<th>Model</th>
<th>$a_0$</th>
<th>$a_1$</th>
<th>$\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filler and Lepage* [2]</td>
<td>0</td>
<td>$10^{-182}$</td>
<td>1.123</td>
</tr>
<tr>
<td>Hoek et al. [3]</td>
<td>-4.32</td>
<td>80.35</td>
<td>1</td>
</tr>
<tr>
<td>Larsson et al. [5]</td>
<td>0</td>
<td>77.24</td>
<td>1.2623</td>
</tr>
<tr>
<td>Le Bricon et al. [9]</td>
<td>4</td>
<td>78</td>
<td>1</td>
</tr>
<tr>
<td>Rule et al. [12]</td>
<td>0</td>
<td>76.6</td>
<td>1.16</td>
</tr>
<tr>
<td>transplant recipients</td>
<td>0</td>
<td>66.8</td>
<td>1.30</td>
</tr>
<tr>
<td>native CKD</td>
<td>0</td>
<td>61.9</td>
<td>1.20</td>
</tr>
</tbody>
</table>

* The Filler and Lepage [2] equation is usually expressed as $\log_{10}(GFR)$.

is especially important since each author can have variants for different conditions, for example Rule et al. [12] report cystatin C-based equations for individuals living with CKD and for kidney transplant recipients (Table I).

The cystatin C-based equations all take the form [33]

$$GFR(c) = a_0 + a_1c^{-\gamma}$$

(3)

where $c$ is the concentration of cystatin C (mg L$^{-1}$) and $a_0$, $a_1$, and $\gamma$ are constants specific to each variant of the equation (Table I).

The serum creatinine-based equations take the form

$$GFR(n,h,r) = \beta_0 + \beta_1 f(y,w,h,g,e)n^{\beta_2}h^{\beta_3}r^{\beta_4},$$

(4)

where $y$, $w$, $h$, $g$, and $e$ are age, weight, height, gender and ethnicity (whether Afro-American for example, respectively). The laboratory measurements are blood urea nitrogen ($n$), albumin ($b$) and serum creatinine ($r$). I assume that $y$, $w$, $g$ and $e$ can be determined accurately, so they contribute no significant error to the estimate of GFR and that $\beta_1 f(y,w,h,g,e)$ can be treated as a constant for particular cohorts. The constants ($\beta_2$) and $\beta_1 f(y,w,h,g,e)$ are specific to each variant of the equation (Table II).

The entries in Tables I and II have been selected from a very large and growing number of equations, each of which appears to have the same form as (3) and (4).

III. UNCERTAINTY OF GFR ESTIMATES

Equations (3) and (4) depend on one or more laboratory measurements which inevitably have uncertainty. Ignoring any other source of error in estimating GFR, the uncertainty in the laboratory data can be used to estimate the uncertainty of the GFR estimates obtained from (3) and (4). This involves the standard Taylor series approach to error analysis [34].

A. Cystatin C-based estimates

From (3) the uncertainty in the GFR estimate due to the uncertainty in the laboratory measurement of cystatin C ($u_c$) is

$$u_{GFR(c)} = \gamma(GFR(c) - a_0)\frac{\mu_c}{c}.$$  

(5)

Here the uncertainty ($u_c$) is the standard deviation ($s$) or variance ($s^2$) of $c$. There are many reports of the coefficients of variation for cystatin C measurements. Some of these are shown in Fig. 1, from which it is clear that $s$ is proportional to the concentration of cystatin C and that the average coefficient of variation (CV) is about 0.046. Since the CV for cystatin measurements is $s/c$, (5) yields

$$s_{GFR(c)} = 0.046\gamma(GFR(c) - a_0),$$

(6)

which is an approximation of the standard deviation of the estimate of GFR due to the error in the estimation of cystatin C alone. For all the GFR equations in Table I, except that of Le Bricon et al. [9], $a_0 \leq 0$ so $s_{GFR(c)} > 0.046GFR(c)$ since $\gamma \geq 1$.

![Fig. 1. Relationship between cystatin C concentration and estimated standard deviation of the measurement ($s$). The solid line is the regression line ($s = (0.006 \pm 0.01) + (0.046 \pm 0.003)c$) and the dashed line represents the 95% confidence band. Estimates of $s$ were obtained from six randomly selected studies in which the cystatin C assay was characterised. : [35]; - [36]; o-](image)

TABLE II

<table>
<thead>
<tr>
<th>Model</th>
<th>$\beta_0$</th>
<th>$\beta_1 f(y,w,h,g,e)$</th>
<th>$\gamma_0$</th>
<th>$\gamma_1$</th>
<th>$\gamma_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cockcroft and Gault* [1]</td>
<td>0</td>
<td>$0.85^{\times}(140 - y) \times w \times 7.13/(72 \times A_{\text{eff}})$</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>MDRD 1 [4]</td>
<td>0</td>
<td>$170 \times 0762^{\times} \times 1.18^{\times}\times 0^{\times}^{170}$</td>
<td>-0.170</td>
<td>0.318</td>
<td>0.999</td>
</tr>
<tr>
<td>MDRD 2 [6]</td>
<td>0</td>
<td>$186 \times 0742^{\times} \times 1.217^{\times}\times 2^{\times}20$</td>
<td>0</td>
<td>0</td>
<td>1.154</td>
</tr>
<tr>
<td>Jelliffe 1 [8]</td>
<td>0</td>
<td>$0.90^{\times}[98 - 0.8 \times (y - 20)]$</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Jelliffe 2 [11]</td>
<td>-6$^d$</td>
<td>$-11^{\times}\times^{1}$</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Salazar-Corcoran* [13]</td>
<td>0</td>
<td>$60^{\times}\times 51^{\times^{1}} \times (137^{\times}1346^{\times} - y) \times$</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>(0.283$^d\times^{1}$</td>
<td>$0.287^{d}$</td>
<td>$\times^{1} \times w \times 10^{\times} 9.74^{d}$</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Rule et al. (native CKD) [12]</td>
<td>0</td>
<td>$273 \times 0.738^{\times} \times 5^{\times}20$</td>
<td>0</td>
<td>0</td>
<td>1.22</td>
</tr>
</tbody>
</table>

Gender is specified by $g = 1$ for females and $g = 0$ for males. Ethnicity is specified by $e = 1$ for Afro-Americans and $e = 0$ otherwise.

* $A_{\text{eff}}$ is given by (2).

$^d$ The factor of $10^d$ is necessary to convert $h$ from cm to m.
The difference between estimates of GFR obtained from two forms of (3) is
\[
\Delta \text{GFR}(c) = \alpha_0 - \alpha'_0 + \alpha_1 c^{r'} - \alpha'_1 c^{-r'}
\]
\[
= \Delta \alpha_0 + \left(1 - \frac{\alpha'_1}{\alpha_1}\right) (\text{GFR}(c) - \alpha_0)
\]
\[ (7) \]
where the second form of (3) is indicated by the prime. The relative error is
\[
\frac{\Delta \text{GFR}(c)}{\text{GFR}(c)} - \Delta \alpha_0 = 1 - \frac{\alpha'_1}{\alpha_1} c^{r-r'}
\]
\[ (8) \]
where 0 \( \leq |r - r'| \leq 0.30 \) and 0.73 \( \leq \alpha'_1/\alpha_1 \leq 1.37 \) based on Table I.

**B. Creatinine-based estimates**

Similarly, the uncertainty in the GFR estimate arising from the uncertainty in the laboratory measurement of serum urea nitrogen (\( u_u \)), albumin (\( u_a \)) and creatinine (\( u_c \)) can be calculated from (4)
\[
\text{uGFR}(r) = \left( \frac{\partial \text{GFR}}{\partial u_u} \right)^2 u_u^2 + \left( \frac{\partial \text{GFR}}{\partial u_a} \right)^2 u_a^2 + \left( \frac{\partial \text{GFR}}{\partial u_c} \right)^2 u_c^2
\]
\[ (9) \]
where \( \partial \text{GFR}/\partial u_u \), \( \partial \text{GFR}/\partial u_a \) and \( \partial \text{GFR}/\partial u_c \) are the appropriate forms of (3).

Of course, if the uncertainties are taken to be standard deviations, \( u_u/n \), \( u_a/b \) and \( u_c/r \) are just the CVs for serum urea nitrogen, albumin and creatinine, which are about 0.038 \[41\], 0.029 \[41\] and 0.028 (Fig. 2), respectively. The CV for nitrogen, albumin and creatinine, which are about 0.038 \[41\], \( \leq 0.033 \), so it has been retained to provide a more conservative estimate of the error. Since \( |\gamma' - \gamma| \leq 0.221 \) and it is more difficult to set bounds on \( p'/p \) because of its complexity (Table II).

**Fig. 2.** Relationship between creatinine concentration and estimated standard deviation of the measurement (\( s \)). The solid line is the regression line \( (s = -0.007 \pm 0.003) \times [\text{creatinine}], r^2 = 0.98 \) and the dashed lines represent the 95% confidence band. Estimates of \( s \) were obtained from six randomly selected studies in which the creatinine assay was characterised: \( \bullet \) [43]; \( \cdot \) [38]; \( \circ \) [44]; \( \alpha \) [39]; \( \nabla \) [24]; \( \square \) [21]. The highest datum does strongly influence the regression, but it is included to provide a conservative estimate of the CV as described in the text.

**IV. RESULTS**

**A. Cystatin C-based estimates**

Glomerular filtration rate declines with increasing \( c \) (Fig 3A) and all of the forms of (3) listed in Table I are indistinguishable at low \( c \) (Fig 3B), as I have shown elsewhere \[33\] in a different way. However, as \( c \) increases, the Filler and Lepage \[2\] and the transplant recipient form of the Rule et al. \[12\] equations become statistically distinguishable from the others (these are the upper two curves in Fig. 3, A and B).

**B. Creatinine-based estimates**

Glomerular filtration rate declines with increasing \( r \) (Fig 4A) and all of the forms of (4) listed in Table II except the Salazar-Corcoran expression \[13\] are indistinguishable at moderate \( r \) (Fig 4B). However, as \( r \) decreases or increases, it is clear that some forms of (4) become statistically distinguishable (Fig. 4, A and B).

**C. The creatinine-cystatin C correlation**

The variability in the correlation between \( c \) and \( r \) \[22-26\] can be rationalised by equating (3) and (4) to obtain
\[
c = \left( \frac{\alpha_0 r}{\beta_0} \right)^{p'}/p
\]
\[ (13) \]
which can be a concave or convex function depending on the parameter values. For some, but not all, forms of (3) and (4), it is clear from Tables I and II that \( \alpha_0 = \beta_0 = 0 \), in which case (13) becomes.
The variability in the correlation between the concentrations of cystatin C and creatinine [22-26] is qualitatively consistent with (13). Moreover, the data reported by Yashiro et al. [26] are at least partly quantitatively consistent with (13). For example, comparison of (14) with their regression data indicates that \( \gamma/\gamma = 0.799 \), which is within the range of \( \gamma/\gamma \in [0.768, 1.22] \) calculable from Tables I and II. The same comparison yields \( (a_1/b)(f(\cdot))^{\gamma/\gamma} = 1.28 \), which could be consistent with the accessible range of \( a_1/b, f(\cdot) \) depending on the combination of age, weight and height of the individual (Table II) and given \( \gamma \geq 1 \) (Table I).

It has been argued that serum creatinine is a poor early indicator of kidney dysfunction [19, 20] and that GFR can decline by half before any significant increase in serum creatinine is observed. However, the sensitivity of GFR to \( c \) or \( r \) can be estimated from (3) and (4), respectively as

\[
c = \left( \frac{\alpha_1}{\beta_0 - \alpha_0} \right)^{1/y},
\]

which also applies at small values of \( r \), if \( \beta_0 - \alpha_0 > 0 \). In either of these circumstances, the correlation between \( c \) and \( r \) will be high. At high \( r \), (13) asymptotically approaches

\[
c \approx \left( \frac{\alpha_1}{\beta_0 - \alpha_0} \right)^{1/y},
\]

in which case the correlation coefficient would tend towards zero. So, one implication of (13) is that the correlation coefficient between \( c \) and \( r \) can be zero at high \( r \) and approach 1 at low \( r \). A second implication is that variability in the strength of the relationship between \( c \) and \( r \) is to be expected.

V. DISCUSSION

Small errors are inherent in the measurements of creatinine and cystatin C (Figs 1 and 2), but they tend to be neglected when estimating GFR and are often not even reported [45]. They can be sufficient to make it difficult to distinguish the cystatin C-based expressions (3) or some of the creatinine-based expressions (4) used to calculate GFR (Figs 3 and 4).

This need not invalidate statistical comparison of the equations [3, 27-32], but does prompt the suggestion the confounding effects of the measurement error should be considered.
\[ \frac{\partial \ln GFR(c)}{\partial \ln c} = -\gamma \frac{GFR(c) - a_0}{GFR(c)} \approx -\gamma \]  
(16)

and

\[ \frac{\partial \ln GFR(r)}{\partial \ln r} = -\gamma \frac{GFR(r) - b_0}{GFR(r)} \approx -\gamma r. \]
(17)

Since \( \gamma \approx \gamma \) (Tables I and II), this indicates that the cystatin C-based and creatinine-based GFR expressions should be similarly sensitive. However, the variation in \( c \) with age and gender is less pronounced than that reported for \( r \) in at least some cases [23], which may contribute to the assessment of the sensitivity of GFR.

In the identification of the best means of estimating GFR in clinical practice, analyte measurement error should be considered. Moreover further analysis of the relationship between the concentrations of cystatin C and creatinine (13) is warranted.

REFERENCES


