Estimated glomerular filtration rate as an end point in kidney transplant trial: where do we stand?

Christophe Mariat, Nicolas Maillard, Manolie Phayphet, Lise Thibaudin, Sylvie Laporte, Eric Alamartine and François Berthoux

Service de Néphrologie, Dialyse et Transplantation Rénale, Laboratoire d’Explorations Fonctionnelles Rénales, EA 3065, CHU de Saint-Etienne, Université Jean Monnet, Saint-Etienne, France

Keywords: clinical trial; cystatin; glomerular filtration rate; predicting equations; transplantation

Introduction

Over the past two decades, substantial improvements in short-term kidney transplant outcome have greatly limited our ability to assess the efficacy of newer therapeutic strategies according to conventional short-term end points, namely the 1-year graft/patient survival and the 1-year acute rejection rate. For instance, with current figures, of less than 15% of patients experiencing an acute rejection episode over the first year post-transplant, it has become logistically challenging (if not clinically questionable) to test a new treatment on its ability to further prevent acute rejection (Table 1). Meanwhile, these traditional end points have failed to predict long-term survival, which is at present a major challenge in renal transplantation. Identification of new, short-term end points capable of correlating with long-term graft outcome is thus necessary. Among other candidates, post-transplant graft function seems to be an attractive alternative end point for clinical trials [1], even though its use as a true surrogate marker for allograft loss has been disputed [2,3]. Comparison of renal graft function is already largely reported as the primary criterion for the evaluation of novel immunosuppressive strategies [4,5].

Of course, before qualifying graft function as a valid end point, one must first make sure to use a valid measure of renal function. The glomerular filtration rate (GFR) is considered as the best overall index of renal function and inulin clearance, the gold standard for its measurement [6]. Other exogenous markers have been advocated for a direct GFR determination, such as radiolabeled isotopes (51Cr EDTA, 99mTc DTPA or 125I Iothalamate) and non-radioactive contrast agents (Iothalamate or Iohexol). Although definitive evidence of their strict equivalence against the standard inulin may be still lacking, these markers are also traditionally deemed as reference methods of GFR measurement. Unfortunately, all these methods, including inulin clearance, are for different reasons not easy to implement. As an alternative, a number of easy-to-use mathematical equations, incorporating different anthropometrical variables in addition to biological parameters, have been developed to predict (‘estimated GFR’), rather than to directly measure GFR (‘true GFR’). These GFR-predicting equations have been widely used to provide a bedside assessment of renal function. While they only give an approximation of true GFR, they allow us to circumvent the various limitations of the use of serum creatinine alone, and certainly offer a more valuable tool to predict GFR in the clinical setting [7]. As a direct corollary, these equations have become very popular in the research setting as well. In transplant trials, they now tend to literally substitute for the reference methods of GFR measurement. However, data have recently been accumulated pointing to some serious limitations of these equations in giving an accurate evaluation of renal graft function. In this report, we review the recent literature that has led to question the performance of these equations when applied to transplant patients.

GFR predicting equations: potential candidates for assessing renal graft function

A great number of mathematical equations have been developed over the years in order to provide physicians with the
The validity of the Cockcroft and Gault formula has been extensively evaluated in different patient populations. However, until recently, very few studies have specifically assessed its ability to assess renal graft function [13].

The Walser equation was first published in 1993 and, similar to the Cockcroft and Gault formula, includes only age, serum creatinine and weight, making it very attractive for a bedside estimate. Unlike the Cockcroft and Gault formula, which gives an estimate of creatinine clearance, which in turn gives an approximation of GFR, the Walser equation provides a direct prediction of GFR. Indeed, this equation was developed in comparison to urinary clearance of $^{99m}$Tc DTPA from a cohort of 85 patients with moderate to severe chronic renal failure (serum creatinine concentration $\geq 177 \mu$mol/l).

The Nankivell formula is the only one that has been computed from a renal transplant population (against a direct measure of GFR by plasma $^{99m}$Tc DTPA clearance). For this reason, this equation has always been seen as more appropriate than any other for assessing renal graft function and thus has been qualified for clinical research. As a matter of fact, the Nankivell formula was integrated in the methods of large international trials, long before the first studies trying to confirm the initial promising data arose.

Levey and colleagues derived different predictive equations from the 1628 patients included in the modification of diet in renal disease (MDRD) study and undergoing renal clearance of $^{125I}$Iothalamate. Since their publication in 1999, the MDRD equations have been presented as a somewhat new standard in GFR prediction, and a considerable number of studies have looked at the added value these equations may actually provide [14–20].

Of note, there is some heterogeneity between the equations as to whether the estimated GFR value needs to be indexed to body surface area (BSA). BSA indexing is not necessary for the MDRD equation (BSA correction was already applied to the reference method this equation was validated against); conversely, BSA indexing is recommended for both the Cockcroft–Gault and Nankivell equations. The rationale for this recommendation is however questionable, since BSA correction has actually little impact on the results in the normal weight population and might even be misleading when GFR is repetitively calculated in longitudinal studies as suggested by Delanaye et al. [21]. This issue is even more complicated for the Walser equation, which was validated against a reference method normalized not to 1.73 m$^2$ of BSA but to 3 m$^2$ of height$^2$.

**How is predictive performance evaluated? How should it be?**

The methodology used to assess the performance of a given GFR-predicting equation is far from being standardized and remains quite heterogeneous across studies. This severely limits the ability to compare the results and obviously contributes to blur the picture of the prediction’s real value. In addition, statistical methods used to analyse the equations’ validity may sometimes be inappropriate from the standpoint of clinical research. In this particular situation, the key question that should be addressed is whether estimated GFR can safely substitute for GFR measured by reference methods.

‘Correlation’ analysis is largely used in studies testing a predictive equation against a reference method and is certainly the most accessible criterion of performance. The correlation coefficient $r$ measures the strength of the relationship between estimated and true GFR and more precisely, the analysis of $r$ indicates how much variability of estimated GFR account for variability of true GFR. However, a significant correlation only means that the null
hypothesis of no relationship between the GFR-predicting equation and the reference method can be rejected. We clearly need more information to assess performance of a GFR estimate.

In 2002, the National Kidney Foundation released clinical guidelines on the evaluation of chronic kidney disease (CKD) [12] and proposed a methodological framework to evaluate GFR-predicting equations according to ‘bias’, ‘precision’ and ‘accuracy’. ‘Bias’ expresses the systematic deviation from the gold standard measure of GFR and is given by the difference between the true and estimated values of GFR (absolute bias). The deviation from the gold standard can also be expressed as a relative difference, i.e. percent deviation from the true GFR (relative bias). This has the advantage of allowing a more valuable evaluation throughout the whole spectrum of kidney function. Clinically this is relevant, as there is less concern about the difference between 100 and 130 ml/min/1.73 m² than biologically this is relevant, as there is less concern about the difference between 30 and 60 ml/min/1.73 m². ‘Precision’ expresses the variability (or dispersion) of predictions around the true GFR and corresponds to the standard deviation of the difference between the true and estimated GFR. ‘Accuracy’ combines precision and bias and is measured by the proportion of estimates falling within a certain percent of the true GFR (e.g. 30% accuracy is the proportion of predicted GFR within ±30% of the true GFR).

‘Bias’, ‘precision’ and ‘accuracy’, as defined by the National Kidney Foundation, are simple and reproducible criteria. They seem to be appropriate to give a fair picture of the predictive performance and naturally tend to be more and more used in studies looking at the validity of GFR-estimating equations.

‘Agreement’, which is best evaluated by the Bland and Altman method, should definitely be part of a standardized evaluation of the predictive performance as well [22]. Indeed, agreement directly indicates how well the equation will substitute for the reference standard itself. Thereby, it is probably the most informative way to describe the predictive performance and to make sure that the equation may safely replace the reference method in the specific setting of clinical trials.

Performance of Cockcroft–Gault, Walser, Nankivell and MDRD equations in predicting renal graft function

Cockcroft–Gault, Walser, Nankivell and MDRD equations are constantly found to closely correlate with measured GFR [23–28]. However, they all displayed considerable lack of agreement with limits of agreement over 35 ml/min/1.73 m² apart [24,25].

Tables 2 and 3 summarize the different levels of accuracy of Cockcroft–Gault, Walser, Nankivell and MDRD equations, as recently reported in different cohorts of renal transplant patients. In keeping with the poor agreement noted between the predicting equations and the direct measurement of GFR, at least 64% of estimated GFR are found to differ from true GFR by ±10% (Table 2).

Table 2. Comparison across studies of the 10% accuracy of MDRD, Walser, Cockcroft–Gault and Nankivell equations in renal transplant patients

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MDRD*</td>
<td>30%</td>
<td>44%</td>
<td>25%</td>
<td>38%</td>
</tr>
<tr>
<td>Walser</td>
<td>28%</td>
<td>46%</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Cockcroft</td>
<td>24%</td>
<td>31%</td>
<td>9.5%</td>
<td>36%</td>
</tr>
<tr>
<td>Nankivell</td>
<td>23%</td>
<td>27%</td>
<td>NA</td>
<td>35%</td>
</tr>
</tbody>
</table>

For each study, the percentage of predicted GFR falling within 10% of the true GFR is reported.

*Equation 7 was used by Mariat et al., the abbreviated version by the others.

Cohort of 294 patients; renal inulin clearance used as the reference method of GFR measurement (mean measured GFR = 49 ml/min/1.73 m²; range: 8–122 ml/min/1.73 m²).

Cohort of 81 patients; plasma iohexol clearance used as the reference method of GFR measurement (mean measured GFR = 56.1 ml/min/1.73 m²; range: 21.8–86.1 ml/min/1.73 m²).

Cohort of 95 patients; plasma clearance of 99mTc-DTPA used as the reference method of GFR measurement (mean measured GFR = 36 ml/min/1.73 m²; range: 11.8–71.0 ml/min/1.73 m²).

Cohort of 798 patients; renal clearance of 125I-iothalamate used as the reference method of GFR measurement (mean measured GFR = 55 ml/min/1.73 m²; range: 18–115 ml/min/1.73 m²).

NA: not available.

Table 3. Comparison across studies of the 20% accuracy of MDRD, Walser, Cockcroft–Gault, and Nankivell equations in renal transplant patients

<table>
<thead>
<tr>
<th></th>
<th>Mariat et al.</th>
<th>Gaspari et al.</th>
<th>Poggio et al.</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDRD*</td>
<td>57%</td>
<td>76%</td>
<td>53%</td>
</tr>
<tr>
<td>Walser</td>
<td>53%</td>
<td>80%</td>
<td>NA</td>
</tr>
<tr>
<td>Cockcroft</td>
<td>46%</td>
<td>57%</td>
<td>37%</td>
</tr>
<tr>
<td>Nankivell</td>
<td>43%</td>
<td>50%</td>
<td>38%</td>
</tr>
</tbody>
</table>

For each study, the percentage of predicted GFR falling within 20% of the true GFR is reported.

* Equation 7 was used by Mariat et al., the abbreviated version by the others.

Cohort of 294 patients; renal inulin clearance used as the reference method of GFR measurement (mean measured GFR = 49 ml/min/1.73 m²; range: 8–122 ml/min/1.73 m²).

Cohort of 81 patients; plasma iohexol clearance used as the reference method of GFR measurement (mean measured GFR = 56.1 ml/min/1.73 m²; range: 21.8–86.1 ml/min/1.73 m²).

Cohort of 209 patients; renal clearance of 125I iothalamate used as the reference method of GFR measurement (mean measured GFR = 44 ml/min/1.73 m²; 10th–90th percentile: 12–80 ml/min/1.73 m²).

Serum creatinine values from the study of Poggio et al. were calibrated with the MDRD laboratory.

NA: not available.
Overall, none of these equations appears to demonstrate a level of agreement with a reference method compatible with clinical research requirements. This is particularly true for clinical trials investigating a therapeutic strategy designed to improve graft function and which aim to compare at one point the mean predicted GFR between two groups. In such trials, it may appear relevant to detect an improvement of 10 ml/min in GFR. It is clearly not if GFR is estimated with one of these equations, since more than one third of the predicted values spontaneously differ from true GFR by more than 10 ml/min [24]. Likewise, testing a hypothesis of superiority of 20% (‘mean GFR under treatment B of 72 ml/min against 60 ml/min under treatment A’, for instance) can seem reasonable, not by using an equation that falsely gives a value beyond this threshold for at least 20% of the estimated GFR (Table 3).

Of note, the superiority of the MDRD equations over at least the Cockcroft–Gault formula is regularly pointed in many recent studies [25–27,29,30]. Similarly, we and others have reported a better predictive performance of the Walser equation [23–25,30]. However, the level of agreement with a reference method for both MDRD and Walser equations still falls far short of what should be required within the context of a clinical trial.

Surprisingly, the Nankivell equation is never found to display a better GFR estimation and more often offers the worst prediction in renal transplant patients. There is no clear explanation regarding the rather poor performance of the Nankivell equation, but this finding, that consistently comes up from one study to another, does not plead for the use of this equation to assess renal graft function, especially in clinical trials.

The 2002 K/DOQI recommendations [12] have adopted a classification into five stages of CKD according to the level of renal function, and have recommended, for the evaluation of GFR, to preferentially use the MDRD study and Cockcroft–Gault equations. These recommendations stem from a systematic analysis of the performance of MDRD and Cockcroft–Gault equations, based on their respective ‘bias’, ‘precision’ and ‘accuracy’ in non-transplant patients. By using the very same analytical methodology, we found that the MDRD equations (equation 7 and the abbreviated MDRD equation) and Cockcroft–Gault formula failed to efficiently categorize transplant patients into the different stages of CKD. Less than two-thirds of the transplant patients are well classified, and are so regardless of the GFR estimates considered [31].

Interestingly, the K/DOQI guidelines based their recommendation in non-transplant patients, on a 30% accuracy reaching 90% for the MDRD equations. In renal transplant patients, we noted a 30% accuracy between 69 and 77% for the four-variable version of this equation. The figure is even worse for the Cockcroft–Gault equation, since only 59% of the estimates fall within 30% of the measured GFR [31]. Raju et al. [28] have reported comparable data from a cohort of 81 renal transplant patients, with a 30% accuracy around 70% for the Cockcroft–Gault and MDRD equations.

More recently, Gera et al. sought to determine whether the Cockcroft and MDRD equations were capable of accurately representing changes in graft function over time [32]. Consistent with previous studies reporting substantial variability of the equations performance at different time points post-transplant [25,27], they found that the predicting equations were less accurate within the first year of transplantation and that they were quite limited in assessing GFR decline over time. Among patients losing GFR at a rate faster than −1 ml/min/1.73 m²/year, only 50% were correctly identified by the MDRD slope as losing graft function.

Are we doomed to exclusively use reference methods to assess renal graft function in clinical trials?

Despite the considerable number of formulas that have been developed over the years, it seems that none of them (including the ones that are currently used in clinical trials) can boast an acceptable predictive performance to substitute for a reference method of direct GFR measurement, at least in situations where an accurate assessment of renal function is required. Therefore, one may ask whether there is still some room to design an nth equation. To answer this question we may need to first figure out why GFR prediction is so challenging in transplantation.

There are probably several potential reasons. First, the majority of these equations has been initially developed from non-transplant patients and does not include many factors specific to transplant recipients that may impact on their predictive performance. For instance, variables such as the number of acute rejections, length of time spent on dialysis or cumulative steroid dose have been shown to be predictors of muscle mass index in transplant recipients, independent of body weight and other parameters usually included in the GFR estimates [10]. Second, the nephron mass transplanted to the recipient is never taken into account and yet is very likely to directly influence the GFR measured after transplantation. So far, even in the equations specifically derived from a transplant cohort, the only anthropometrical indices that are computed come from the recipients. Since the renal mass is certainly correlated to the body habitus, some corrective variables relative to the donor characteristics might be required to enhance the performance of these estimates in the transplant setting. In this respect, it could be worth computing a new, transplantation-specific GFR-predicting equation, incorporating some donor parameters.

However, one has to keep in mind the fact that other limiting factors exist and further jeopardize the reliability of these equations when applied not only to transplant patients, but also to the general population. Among them, the patients’ demographics are likely to affect the validity of the equation, and to some extent the predictive performance is closely dependent on the initial study population characteristics. For example, while the MDRD study equations have been developed and validated in a population exhibiting a mean GFR of 39.8 ml/min/1.73, several reports have found that their performance can be dramatically hampered in individuals with mild decrease in kidney function or normal GFR [20]. Similarly, the choice and calibration of the assay used to determine the serum creatinine concentration has been pinpointed as a potential misleading factor [7,33,34].
and has recently led to re-express the MDRD equation for use with a standardized creatinine assay [35,36]. Whether this effort of standardization will translate into a better prediction for transplant patients is still unclear. Indeed, the calibration-related error is particularly significant for the normal range of creatinine concentrations, which is not the typical range of creatinine values seen in transplantation. In a cohort of 209 renal transplant patients with a mean serum creatinine of 2.4 mg/dl, Poggio et al. did not report a better prediction of the MDRD equation, despite careful calibration of serum creatinine values with the MDRD laboratory [29].

Another possible way to improve the predictive performance of the existing equations is to block the tubular secretion of creatinine with an oral administration of cimetidine. The tubular secretion of creatinine increases with the degree of renal impairment and is likely to account at least partially for the overestimation of the prediction observed in transplant patients. Many investigators have reported that the predictive performance of cimetidine-corrected creatinine clearance and cimetidine-corrected Cockcroft–Gault formula substantially improved in patients with CKD [37–39]. Kemperman et al. concluded that in renal allograft recipients, the accuracy of the Cockcroft–Gault formula 24 h after 2400 mg of cimetidine was also significantly enhanced and proposed its use for routine follow-up of transplant patients [40]. A recent study sought to extend the use of cimetidine-corrected creatinine values to the MDRD equation [41]. However, the exact performance of this so-corrected equation is still an open question, since the predicted GFR in this study was not compared to the true GFR but to creatinine clearance.

Cystatin C is a low-molecular mass protein acting as a protease inhibitor and produced in a constant manner by all nucleated cells. It is freely filtered by the glomerule and catabolized in the proximal tubule without being secreted. Thus, its plasmatic concentration can be used to estimate GFR [42]. Actually, serum cystatin C has long been considered at least equal if not superior to serum creatinine as a marker of GFR [43]. Two recent studies have compared the predictive performance of serum creatinine and cystatin C-based equations in renal transplant patients [44,45]. They both found that cystatin C equations perform better than traditional creatinine-based equations. However, the potential advantages of cystatin C-based equations over creatinine-based equations in transplant patients have not been confirmed in a more recent study [46]. This apparent discrepancy will certainly stimulate further research to better understand the different factors other than GFR that are likely to affect cystatin C levels. Among them, the impact of calibration of the different assays used to measure cystatin C as well as the exact influence of immunosuppressive drugs (as steroid) will definitely have to be addressed before qualifying any cystatin C equation in transplantation.

**Conclusion**

Renal graft function is now a key criterion for evaluating new therapeutic strategies in kidney transplantation. Using a precise and reliable method to measure an end point in clinical trials is a prerequisite to avoid flawed and misleading interpretation. In this context, the GFR-predicting equations that have been recommended so far cannot guarantee an accurate assessment of renal graft function. The recent literature on kidney transplantation does not support the widespread use of any of these equations for clinical research. At present and until reliable alternatives are fully validated, reference methods of direct GFR measurement should be systematically considered when designing a kidney transplant trial.

**Acknowledgements.** We thank Aaron Greenman for his help in editing the final version of the manuscript.

**Conflict of interest statement.** All the authors declare that they have had no involvement that might raise the question of bias in the work reported or in the conclusions, implications or opinions stated.

**References**


35. Vickery S, Stevens PE, Dalton RN et al. Does the ID-MS traceable MDRD equation work and is it suitable for use with compensated Jaffe and enzymatic creatinine assays? Nephrol Dial Transplant 2006; 21: 2439


Received for publication: 17.7.07
Accepted in revised form: 11.9.07