SYSTEMATIC REVIEW

Methods to evaluate renal function in elderly patients: a systematic literature review

Gijts Van Pottelbergh1,2,3, Lore Van Heden1, Catharina Mathei1,2, Jan Degryse1,2

1Department of Primary Health Care, Katholieke Universiteit Leuven, Leuven, Belgium
2Institute of Health and Society, Université Catholique de Louvain, Brussels, Belgium
3Fellow of the Research Foundation Flanders, Belgium

Address correspondence to: G. Van Pottelbergh. Tel: +32 1633 7480; Fax: +32 1633 7480. Email: gijts.vanpottelbergh@med.kuleuven.be

Abstract

Context: multiple studies of elderly patients show that the prevalence of chronic renal failure in people aged 65 years and older is dependent on the method used to calculate the glomerular filtration rate. We performed a systematic literature search with research question: What is the best method that could be applicable in clinical practice for evaluating renal function in the elderly? Studies using inulin, Cr-51-EDTA, Tc-DTPA or iohexol assays as the gold standard were included.

Methods: we searched the PubMed and EMBASE databases. Articles found were screened first by title and abstract and then by five criteria. Retained articles were scored using an adapted version of QUADAS.

Results: twelve articles had an identified population or subpopulation aged 65 years and older. The studies were heterogeneous with regard to the population investigated and the statistical procedures used to compare the methods and equations with the gold standard. The Cockcroft–Gault (CG) and MDRD equations and the serum cystatin C concentration produced the highest correlations with the gold standard.

Conclusions: no accurate method to evaluate renal function in the elderly was found. Serum cystatin C concentration and the CG and MDRD formula might be valuable parameters, although there is insufficient evidence.

Keywords: chronic kidney disease, Cockcroft–Gault formula, elderly patients, kidney function tests, MDRD formula, systematic review

Introduction

Chronic kidney disease (CKD) is recognised as an important problem in public health for several reasons. First, the pathology has a high prevalence worldwide [1], and the prevalence of impaired renal function increases with advancing age. Given the demographic evolution of the population of Europe, it is probable that the total prevalence of CKD will increase markedly in the coming years.

Secondly, CKD is an independent cardiovascular risk factor: a low glomerular filtration rate (GFR) is associated with increasing mortality, more cardiovascular events and more hospitalisations [2].

The use of gold standard tests such as inulin clearance, Cr-51-EDTA or Tc-DTPA is not possible in a routine clinical practice, and the Cockcroft–Gault (CG) formula [3] was developed as an alternative. Since then, many researchers have tried to find a better formula. In 1999, Levey et al. [4] published their formula based on a large study, the MDRD Study Group. Renal function can be estimated based on serum creatinine concentration and age. Since then, many studies have compared the CG and MDRD formulas and with other equations [5,6], both with and without reference to a gold standard. In recent years, cystatin C levels have also been used to determine the GFR. However, none of these methods has been validated in a large population of elderly patients, leading to the following question: What is the best method, applicable in ambulatory practice, to evaluate kidney function in the elderly?

To our knowledge, no systematic review has investigated which formula produces the best measure of renal function in a population of patients aged 65 years and older.

Search strategy and selection criteria

A ‘PIRT’ (Patient—Index test—Reference test—Target condition) [7], analogous to the ‘PICO’ [8] (Patient—Intervention—
Table 1. Description of the 12 studies with an elderly (sub) population

<table>
<thead>
<tr>
<th>Year</th>
<th>Design</th>
<th>Standard</th>
<th>Mean GFR</th>
<th>SD</th>
<th>n total</th>
<th>n aged &gt;65 years</th>
<th>Age range</th>
<th>Location</th>
<th>Method for recruiting patients</th>
<th>Creatinine assay</th>
<th>Cockcroft–Gault</th>
<th>MDRD</th>
<th>Cyst C</th>
<th>Creat Cl</th>
<th>Other formulas</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td>Prospective</td>
<td>DTPA</td>
<td>79.4</td>
<td>17.1</td>
<td>850</td>
<td>134</td>
<td>65–93</td>
<td>Montpellier</td>
<td>Referred for GFR measurement</td>
<td>Jaffé enzymatic</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>2005</td>
<td>Prospective</td>
<td>Inulin</td>
<td>90</td>
<td>39.1</td>
<td>380</td>
<td>35</td>
<td>65–88</td>
<td>Naples</td>
<td>Volunteers</td>
<td>Jaffé</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>2005</td>
<td>Prospective</td>
<td>Inulin</td>
<td>93.6</td>
<td>22.9</td>
<td>61</td>
<td>61</td>
<td>62–91</td>
<td>Mannheim</td>
<td>Acute admission in hospital</td>
<td>Jaffé</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>2005</td>
<td>Retrospective</td>
<td>EDTA</td>
<td>43.3</td>
<td>26</td>
<td>2,095</td>
<td>30</td>
<td>68–76</td>
<td>Paris</td>
<td>Referred for GFR measurement</td>
<td>Jaffé enzymatic</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>2003</td>
<td>Prospective</td>
<td>Inulin</td>
<td>51</td>
<td>18.4</td>
<td>30</td>
<td>30</td>
<td>69–92</td>
<td>Ireland</td>
<td>Ambulatory practice in various hospitals</td>
<td>Jaffé</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>2002</td>
<td>Prospective</td>
<td>EDTA</td>
<td>83.3</td>
<td>41.7</td>
<td>53</td>
<td>41</td>
<td>57–90</td>
<td>Hanover</td>
<td>Geriatric ward</td>
<td>Jaffé</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>2002</td>
<td>Prospective</td>
<td>EDTA</td>
<td>104</td>
<td>12</td>
<td>420</td>
<td>222</td>
<td>67 ± 6 SD</td>
<td>Heidelberg</td>
<td>Members of Academy for Elderly or via the renal ward</td>
<td>Jaffé enzymatic</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>2001</td>
<td>Prospective</td>
<td>Inulin</td>
<td>60.5</td>
<td>30.9</td>
<td>18</td>
<td>18</td>
<td>65–88</td>
<td>Philadelphia</td>
<td>At coronary angiography</td>
<td>Jaffé</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>1998</td>
<td>Prospective</td>
<td>Iohexol</td>
<td>54.7</td>
<td>11.99</td>
<td>46</td>
<td>46</td>
<td>66–82</td>
<td>New Zealand</td>
<td>Seniors club</td>
<td>Jaffé</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>1991</td>
<td>Prospective</td>
<td>DTPA</td>
<td>58.2</td>
<td>28.8</td>
<td>48</td>
<td>48</td>
<td>69–92</td>
<td>Ghent</td>
<td>Acute admission to hospital</td>
<td>Jaffé</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>2005</td>
<td>Prospective</td>
<td>EDTA</td>
<td>53.3</td>
<td>17.8</td>
<td>52</td>
<td>52</td>
<td>69–92</td>
<td>Kent</td>
<td>Ambulatory practice</td>
<td>Jaffé</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

IDMS, isotope dilution mass spectrometry.
Table 2. Regression analysis of the calculated GFR in comparison with the gold standard

<table>
<thead>
<tr>
<th></th>
<th>O’Riordan et al. [16]</th>
<th>Lamb et al. [17]</th>
<th>Van den Noortgate et al. [26]</th>
<th>Lamb et al. [27]</th>
<th>Spinler et al. [18]</th>
</tr>
</thead>
<tbody>
<tr>
<td>CG</td>
<td>0.79</td>
<td>0.69</td>
<td>0.82</td>
<td>0.84</td>
<td>0.75</td>
</tr>
<tr>
<td>MDRD</td>
<td>/</td>
<td>0.79</td>
<td>0.65</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Creatinine clearance</td>
<td>0.62</td>
<td>/</td>
<td>0.57</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Serum creatinine</td>
<td>0.64</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Cystatin C</td>
<td>0.79</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
</tbody>
</table>

Four studies reported $R^2$ values for evaluating the various methods to estimate renal function in comparison with the gold standard. The report by Spinler et al. indicated an $R$ value. / means no data available in study.

Results

Using the search terms, 758 articles were detected in PubMed and 58 in EMBASE. After combining these two searches, 34 references appeared to be redundant, although 14 new references were found by searching EMBASE. Of these 772 articles, the first reviewer retained 344 articles after the first screening round and the second reviewer 376. Altogether, the remaining 397 articles were screened a second time using the five criteria; 123 articles were retained and were assessed based on their quality, and 77 articles scored 8 or more on the 11-point scale. Backward tracking resulted in 42 articles that seemed relevant based on their titles; 10 of these articles met the five criteria and were of good quality. We retained 87 articles.

It was impossible to pool all the data for patients aged 65 years and older because no separate information was presented for the elderly patients included. Hence, we limited the data to those obtained from articles that either included only an elderly population or described a separate subpopulation of patients aged 65 years and older. This was the case in 14 articles. Two articles were excluded at this point. One investigated the influence of NSAIDs on the renal function in elderly people [14], but did not compare the gold standard with formulas. The second article excluded [15] also studied the influence of medication and used data from another study, which had already been included. Of the 12 articles remaining, two [16, 17] used the same data set. Both articles were retained because they used different statistical procedures, but they were matched to each other so that our analysis would not evaluate the same group of patients twice.

Twelve articles [16–18, 18–27], two of which included data on the same group of patients, were selected. Table 1 shows the size, mean GFR values and standard deviations (SDs) of these studies. It is clear that the studies included different ranges of renal function; this is a consequence of the heterogeneity of the populations selected for these studies. For example, Fliser and Ritz [24] measured a mean GFR of 104 ml/min in healthy, active seniors. This value contrasts with that reported by Froissart et al. [22], who selected their patients retrospectively from databases of patients who had been referred for true GFR measurement. Many studies used a population of patients who had been referred to hospital for some reason. Such selection procedures lead to selection...
bias because there has to be a strong reason to measure the GFR by an invasive gold standard method. In Table 1, the gold standard method used to analyse serum creatinine and the methods used to estimate the GFR in the different studies are indicated.

The mean difference between the measured GFR and the calculated GFR was determined in several studies. Figure 1 shows the mean differences in the calculated GFR and the measured GFR and 95% CIs (when given) for the CG and MDRD equations. In all articles, serum creatinine concentration was measured using the Jaffé technique. However, in the two articles that compared the data obtained using both the enzymatic and Jaffé methods to measure serum creatinine concentration, the mean GFR differed between the techniques. Calculating the GFR using the CG and MDRD formulas, Verhave et al. [19] reported GFR values of −11.3 and −5.0 ml/min for the enzymatic method versus −22.4 and −18.0 ml/min for the Jaffé method. Lamb et al. [17] also found a difference in the same direction: −5.9 and 5.7 ml/min for the enzymatic method versus 0.8 and 14.1 ml/min for the Jaffé method. From Figure 1, it is also clear that the method used to measure serum creatinine as well as the gold standard plays a large role in determining the mean difference. The three studies that used EDTA as the gold standard showed similar mean difference values when calculated using the CG and MDRD formulas. These three studies give a slight negative mean difference for values calculated using the CG equation and a slight positive mean difference for values calculated using the MDRD equation. The other three studies used a different gold standard, and their values differed from those of studies using EDTA as the standard. Because these are single studies using a certain gold standard, it is impossible to say whether these differences relate to the study populations or to the gold standard tests used.

Five studies included regression analysis (see Table 2). Spinler et al. [18] calculated and reported the \( R^2 \) value, the other four reports indicated the \( R^2 \) value. Comparing the values obtained after regression analysis, Lamb et al. [27] obtain the same \( R^2 \) value for the MDRD and CG equations com-
pared with the gold standard. In the article by Van Den Noortgate et al. [26], the $R^2$ value was closer to 1.0 for the CG formula than for the MDRD formula. In contrast, in the article by Lamb et al. [17], the $R^2$ value was closer to 1.0 for the MDRD formula. All studies found a lower $R^2$ value for creatinine clearance than the $R^2$ values from the CG and MDRD formulas. O’Riordan et al. [16] found a lower $R^2$ value for serum creatinine concentration than for the CG formula but that the $R^2$ value for cystatin C concentration was the same as that for the CG formula.

From four articles, we could extract data for the construction of a classic 2 × 2 table. These studies used four cut-off values (50, 60, 80 and 90 ml/min) to calculate the sensitivity of the various methods for estimating GFR (MDRD, CG, serum creatinine concentration and creatinine clearance and serum cystatin C concentration). The GFR measured with the gold standard test determined whether the patient had an impaired renal function or not. O’Riordan et al. [16] presented the data as the proportion of positive patients of the total number of patients whose GFR value according to the gold standard test was below the limit of 80, 60 or 50 ml/min, and for GFR calculated using the CG formula, serum creatinine concentration, creatinine clearance and serum cystatin C concentration. Burkhardt et al. [21] presented the data graphically as a receiver-operating characteristic (ROC) plot. We were able to estimate the sensitivity from this graph by choosing a specificity of 80%, and we could convert this ROC curve to absolute numbers to construct a 2 × 2 table for values calculated using the CG and MDRD equations, and using serum creatinine concentration and creatinine clearance, for limits of 60 and 90 ml/min. The article by Van Den Noortgate et al. [26] presented values for the sensitivity and specificity for serum creatinine concentration and creatinine clearance for the limit of 80 ml/min. The article by Lamb et al. [27] presented data for the limit of 50 ml/min, the number of false negatives and false positives for the CG and MDRD formulas and creatinine clearance. Further to these four articles, Nicoll et al. [25] presented a full data set, and we could calculate the four-field tables for this population so we had in total data from five studies. Despite the heterogeneity of these five studies, we pooled all data (in the form of absolute numbers) in one four-field table for the separate studies. Top: Sensitivity of the CG and MDRD formulas, serum creatinine and cystatin C concentrations and creatinine clearance for the pooled data with cut-off values of 50, 60, 80 and 90 ml/min. Middle: Positive predictive value of the CG and MDRD formulas, serum creatinine and cystatin C concentrations and creatinine clearance for the pooled data with cut-off values of 50, 60, 80 and 90 ml/min. Bottom: Sensitivity of the CG and MDRD formulas, serum creatinine and cystatin C concentrations and creatinine clearance for the cut-off value of 60 ml/min for pooled data and for the separate studies. CG, Cockcroft–Gault formula; MDRD, Modification of Diet in Renal Disease formula; Creat S, serum creatinine; Creat CL, creatinine clearance; cyst C, cystatin C.

Figure 2. Mean difference between the gold standard and the formula to calculate the GFR (with 95% CIs) for the separate studies. Top: Sensitivity of the CG and MDRD formulas, serum creatinine and cystatin C concentrations and creatinine clearance for the pooled data with cut-off values of 50, 60, 80 and 90 ml/min. Middle: Positive predictive value of the CG and MDRD formulas, serum creatinine and cystatin C concentrations and creatinine clearance for the pooled data with cut-off values of 50, 60, 80 and 90 ml/min. Bottom: Sensitivity of the CG and MDRD formulas, serum creatinine and cystatin C concentrations and creatinine clearance for the cut-off value of 60 ml/min for pooled data and for the separate studies. CG, Cockcroft–Gault formula; MDRD, Modification of Diet in Renal Disease formula; Creat S, serum creatinine; Creat CL, creatinine clearance; cyst C, cystatin C.
Discussion

The mean difference is not such a good parameter for comparing various methods to a gold standard test because there can be outliers in both positive and negative directions, which can compensate for each other. The results collected in Figure 1 suggest that the choice of gold standard and the method for measuring creatinine concentration and clearance influence the mean difference.

Analysis of the results of the regression analyses leads us to conclude that the MDRD and CG equations correlated more strongly with the gold standard than do serum creatinine concentration and creatinine clearance values. The data from these studies show that the data from the MDRD and CG formulas and the serum cystatin C concentration have similar correlations with the gold standards. However, the cystatin C concentration should be interpreted with caution because this was calculated only from one study of 53 patients.

We compared the various studies in more detail around the cut-off point of 60 ml/min (Figure 2) because this is a clinically relevant value. Below this value, a patient is considered to have renal failure. At this cut-off value, the CG formula seems to score better than the MDRD equation. However, the sensitivity of the MDRD equation around 60 ml/min was calculated from only one study, by Burkhardt et al. [21]. This study also calculated the sensitivity of the CG formula, which produced a significantly lower value. The sensitivity of cystatin C concentration was calculated in only one study, by O’Riordan et al. [16]. In clinical practice, when determining whether a certain patient has renal failure, it is important to know the sensitivity and the predictive value of the test. In clinical practice, it is usually not as important to know the exact GFR if in the normal range, as it is to identify those with low or very low GFR. Thus, it is critical to be able to situate a patient with respect to the cut-off points of 60, 45 and 30 ml/min. It is therefore important for physicians to have a test that performs well especially around these values. Few studies have investigated this question, and the studies we found indicate that both the CG and MDRD formulas score reasonably well. However, in studies with a younger population [22], the CG formula still scores relatively well at 60 ml/min (77.9% true positive rate for GFR < 60 ml/min), but it performs worse at lower values (67.7% true positives rate for GFR < 30 ml/min). In contrast, the MDRD formula scores similarly well at around 60 ml/min (78.1%), but performs better at lower GFR values (78.8%). These lower GFR values were investigated in too few studies to draw firm conclusions. How these formulas score around lower GFR values in the elderly is an essential question that needs to be answered before deciding which formula is best for evaluating renal function in this population. It also is possible that the results from the MDRD formula were not so precise because of the different calibration methods used for the creatinine assay. It had been shown that using the same calibrating method as used in the original MDRD study [28] or using recalibrated IDMS assays [29] gives a better estimation of the GFR when using the MDRD formula. Of the seven studies reporting estimations of the GFR using the MDRD formula, only two (that by Froissart et al. [22], who calibrated their assay with that of the original MDRD study, and that of Lamb et al. [17], who used the IDMS method) reported such a calibration of their creatinine assay.

We had major problems extracting the data from the studies. Not only did we find four gold standards and different creatinine-calibrating methods but also the statistical methods used to analyse the results were very different. The authors did not perform the same type of analyses and even describe different concepts (such as ‘precision’ and ‘accuracy’) with the same name. This limited the possibilities of comparing and pooling the data. Therefore, we decided to report only the mean difference of estimated and measured GFR, regression analysis and sensitivity and specificity for a single reason only: these were the only results we could compare between the different studies. Future investigators should report their statistical analyses in a more standardised way, for instance as suggested in the article by Stevens et al. [30].

Conclusion

We conclude that the serum creatinine concentration is an insensitive measure for evaluating renal function in the elderly because it correlated poorly with the gold standard. Of all methods, the CG and MDRD equations produced the best results for correlation with the gold standard and for sensitivity and specificity.

Based on the available studies, we believe that it is not possible to say which of these two formulas is better for evaluating renal function in people aged 65 years and older. Estimating the GFR based on serum cystatin C concentration has not been studied extensively in the elderly, but it seems a promising method. There is a pressing need for a sound prospective study using standardised creatinine assays, a correct gold standard and a population truly representative of the total population of elderly people.

Key points

- Limited numbers of qualitative studies were found.
- The MDRD and Cockcroft-Gault formula give a more accurate estimation of the GFR in elderly patients.
- Creatinine and urinary creatinine clearance are less accurate methods to estimate the renal function in elderly.
- There is very limited but promising evidence concerning the use of serum cystatin C in elderly patients.

Conflicts of interest

None declared.
Supplementary data

Supplementary data mentioned in the text is available to subscribers in Age and Ageing online.

References


Received 21 October 2009; accepted in revised form 26 May 2010