Original Article

Coronary artery calcification scores in patients with chronic kidney disease prior to dialysis: reliability as a trial outcome measure

Katherine A. Barraclough¹, Lesley A. Stevens², Lee Er¹, Debbie Rosenbaum¹, Jacqueline Brown³, Pari Tiwari³ and Adeera Levin¹

1Division of Nephrology, University of British Columbia, Vancouver, British Columbia, Canada, 2Division of Nephrology, Tufts-New England Medical Center, Boston, MA, USA and 3Department of Radiology, St Paul's Hospital, Vancouver, British Columbia, Canada

Abstract

Background. Coronary artery calcification (CAC) is prevalent in patients with chronic kidney disease (CKD). Data on the reliability and validity of high-resolution computerized tomography (HRCT) in patients with CKD is lacking. The purpose of this study was to evaluate the interand intra-reviewer agreement and inter-scan reproducibility of CACS measurement with HRCT in a cohort of patients with CKD prior to dialysis, and to compare the change in CACS at 30 minutes to the change in CACS over 1 year.

Methods. Thirty-three patients with CKD not yet on dialysis underwent an HRCT scan at baseline and 1 year to assess for CAC and CAC progression. Two radiologists independently reviewed films and each radiologist re-reviewed a randomly selected subset of films they had previously viewed, to assess for inter-reviewer and intra-reviewer reliability, respectively. Patients underwent a repeat scan within 30 min of the first baseline scan to assess for inter-scan reproducibility.

Results. At baseline, eight patients (24%) had no CAC. Of the 25 patients (76%) with CAC, 10 (40%) had severe calcification. Intra-reviewer agreement was 83%. Inter-reviewer agreement ranged between 77 and 94%. Six (27%) of the patients with >30 baseline CACS had >15% change in CACS following repositioning. Four of these patients had an increase in CACS with position change [18% (95% CI: 5–40%)]. Of the 21 patients who underwent a follow-up scan at 1 year, 7 (33%) demonstrated CACS progression. **Conclusions.** There is significant imprecision in HRCTderived CACS in CKD patients. This suggests a need for standardization of methods of CACS measurement with HRCT.

Keywords: cardiovascular disease; chronic kidney disease; coronary artery calcification; high-resolution computerized tomography; test reliability

Introduction

Cardiovascular disease (CVD) is the most frequent cause of morbidity and mortality in patients with chronic kidney disease (CKD) [1]. Risk factors for CVD include traditional cardiac risk factors, as well as non-traditional risk factors such as abnormalities of mineral metabolism. These abnormalities have been associated with mortality and CVD morbidity in epidemiological studies, as well as physiological studies of arterial stiffness and vascular calcification $[2-4]$.

Multiple general population studies now describe the association of coronary artery calcification (CAC) with cardiovascular morbidity and mortality. High-resolution computerized tomography (HRCT) is a highly sensitive technique for detecting CAC [5] and provides a quantitative coronary artery calcification score (CACS).

It has been shown that the magnitude of CAC is substantially higher in patients with CKD than in the general population. Patients on haemodialysis have higher scores compared to those at earlier stages of CKD [6–9]. However, the CACS was developed and validated in the general population without CKD, and data regarding its validity and the significance of CACS levels in the CKD population are scarce. Of greater importance, the reliability of the CACS is not documented in CKD. These data are important with increasing use of the CACS in outcome studies and clinical trials involving CKD patients.

The purpose of this study was to evaluate the inter- and intra-reviewer agreement and inter-scan reproducibility of CACS measurement with HRCT in a cohort of patients with CKD prior to dialysis, and to compare the change in CACS at 30 min to the change in CACS over 1 year. These data on

Correspondence and offprint requests to: Katherine Barraclough, Department of Nephrology, St. Paul's Hospital, 1081 Burrard Street, Providence Building, 6th Floor, Room 6010A, Vancouver V6Z 1Y6, Canada. Tel: +1-604-8754111; Fax: +1-604-8068120; E-mail: arbieb@hotmail.com

^C The Author [2008]. Published by Oxford University Press on behalf of ERA-EDTA. All rights reserved. For Permissions, please e-mail: journals.permissions@oxfordjournals.org

the reliability of the CACS in this patient population will facilitate the interpretation of the test.

Methods

Study design and procedures

Patients. Patients cared for in the Kidney Function Clinic at St. Paul's Hospital, Vancouver, British Columbia, who had an estimated glomerular filtration rate (GFR) between 15 and 60 ml/min/1.73 m² (0.25–1.0 ml/s/1.73 m²) (CKD Stages 3 and 4) were eligible for inclusion. Exclusion criteria included the expected life expectancy of <1 year from non-cardiac causes (e.g. AIDS, active malignancy), acute severe illness, conditions preventing cardiac gating during HRCT (documented heart rate >90 beats/min, atrial fibrillation or flutter), demonstrated progression of kidney dysfunction >5 ml/min/1.73 m² in the previous 6 months, pregnancy or a plan to become pregnant and refusal to provide informed consent. Consecutive patients fulfilling the eligibility criteria were approached and consented, from the existing Kidney Function Clinic population. A total of 33 patients were included.

Clinical data extraction

Medical records were reviewed to obtain demographic data, information on the presence of traditional cardiac risk factors, coronary artery disease and other comorbidities, medications, blood pressure and laboratory results. Current smoking status, the presence of traditional cardiac risk factors, cardiac history and medications were confirmed with the patient. Diabetes was defined as present if the patient was prescribed anti-hyperglycaemic medications, had an elevated haemoglobin A1C or diabetes as a cause of CKD. Hypertension was defined by blood pressure >130/80 or taking antihypertensive medications. Patients were considered to have coronary artery disease if they had a history of previous myocardial infarction or angina, or a positive stress test or angiogram. Estimated GFR was calculated from the serum creatinine measured in closest proximity to the HRCT scan using the four-variable Modification of Diet in Renal Disease (MDRD) Study equation [10].

Imaging procedures

Patients underwent two non-contrast, 64 multi-slice HRCT scans within 30 min of each other, with a period of ambulation between the studies. A further scan was performed at 1 year to evaluate the progression of CAC. All scans were performed in the same room, using the same equipment. HRCT cine acquisition collected 2.5-mm contiguous axial slices from the tracheal carina to the inferior margin of the heart (GE Medical Systems Lightspeed Plus, 120 kV, 10 mA). All areas of calcification with a minimum density of 130 Hounsfield Units within the borders of the coronary arteries were computed. Images were recorded during breath-holding sessions. CT imaging was triggered at 80% of the R-R interval. The acquired images were reviewed on a dedicated workstation (Advantage Work Station, General Electrics). A calcium score was calculated by the Agatston

method [11], where the area of the calcified plaque is multiplied by a weighted coefficient based on the peak density of the calcification, and the CACS is expressed in Agatston units (AU). A CACS was calculated individually for the left main, left circumflex, left anterior descending, posterior descending and right coronary arteries. The scores were then summed to calculate the total CACS.

Scoring of scans

A single expert radiologist reviewed all films. A second expert radiologist reviewed a randomly selected subset of films to determine inter-reviewer agreement. Each radiologist then re-reviewed a subset of films they had previously viewed, with a mean of 7 days between readings. For all readings, the radiologists were blinded to all patients and any prior information on the scan.

Statistical analysis

Descriptive statistics for the continuous variables are presented as mean with standard deviation (SD), or median with inter-quartile range (IQR) depending on distribution. Continuous baseline variables were compared using the one-way ANOVA or Kruskal–Wallis tests as appropriate. Categorical variables were compared using the Fisher's exact test. Calcification data were categorized into CACS groups according to the classification system that has been validated for the quantification of CACS in the general population and modified for use in dialysis populations: no calcification (0 AU), mild to moderate (1–400 AU), severe $(401-1000 \text{ AU})$ and very severe $(>1000 \text{ AU})$ [12–14]. The mean of the scores calculated by the two radiologists, where available, was treated as the score at each time point, with the aim being to minimize error from each reviewer.

Change in CACS after repositioning and 1 year was described in two ways: by difference and percent difference. Change after repositioning was also described by absolute difference and percent absolute difference from baseline. Because of inability to calculate percent change in those with a zero baseline CACS, the percent change from baseline was only applicable to those with $a > 30$ baseline CACS [12,13]. For patients with a CACS $>$ 30, a true change was defined as $a > 15\%$ change from baseline, consistent with previous studies [12,13]. An exact binomial 95% confidence interval (CI) for the proportion of patients with an >15% increase from baseline was constructed. For the intra-reviewer and inter-reviewer reliabilities, we looked at the total score difference within 15% of each other for the two readings/radiologists. All tests were two-sided, with *P*value <0.05 considered significant. Analyses were carried out using SAS software, version 9.1 (SAS Institute, Cary, NC).

The St. Paul's Hospital Ethics Committee approved the protocol. All patients provided written informed consent. The results of the radiological tests were not formally reported to any attending physicians.

Results

Table 1 shows the baseline demographic and clinical characteristics of the cohort according to the CACS grade. The

Reliability of CACS in chronic kidney disease 3201

Table 1. Summary of baseline characteristics

Values expressed are mean [SD], except median [IQR] for estimated GFR, PTH and triglycerides. Estimated GFR was calculated from the MDRD study equation. CAD, coronary artery disease; PTH, parathyroid hormone; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol. To convert haemoglobin in g/dL to g/L, multiply by 10; albumin in g/dL to g/L, multiply by 10; GFR in ml/min/1.73 m² to ml/s/1.73 m², multiply by 0.0167; calcium in mg/dL to mmol/L, multiply by 0.2495; phosphate mg/dL to mmol/L, multiply by 0.3229; PTH in pg/mL to ng/mL, multiply by 1; total cholesterol, HDL-C and LDL-C in mg/dL to mmol/L, multiply by 0.0259; triglycerides in mg/dL to mmol/L, multiply by 0.0113.

mean (standard deviation) age of the cohort was 70 (12) years. Sixteen patients were classified as Stage 3 CKD and 17 as having Stage 4 CKD (Figure 1).

Coronary artery calcification

The median (IQR) CACS at baseline was 379 (23, 1028) (Table 1). CACS was 0 in eight patients (24%), 1–400 in nine patients $(27%)$ and $401–1000$ in six patients $(18%)$. Ten patients (33%) had scores in the severely calcified range (>1000) . The distribution of severity of CACS, using conventional breakpoints, is demonstrated in Figure 1. The patients without any calcification were younger, had a lower pulse pressure and were less likely to have diabetes. Older age, higher pulse pressure, lower albumin and a prior history of coronary artery disease significantly correlated with the severity of baseline CACS. There was no relationship between estimated GFR or CKD stage and CACS (Figure 1).

The CAC scoring results

Inter-reviewer agreement. Table 2 shows inter-reviewer agreement for each of the two baseline scans and for the

1-year follow-up scan. When considering all patients, including those with zero baseline CACS, the median difference in CACS was 0 (−0.5, 36.5). For patients with a baseline score >30, the reviewer agreement ranged between 77 and 94% for the three scan comparisons (Table 2).

Intra-reviewer agreement. For the intra-reviewer component, the median difference (IQR) between scoring results was 0 (−1.5, 4.5). Intra-reviewer comparisons were made on scans of six patients with a baseline CACS >30. For this subset, there was an agreement of CACS scores within 15% of each other (Table 2) for five out of six (83%) scans.

Inter-scan reproducibility. The median absolute difference (IQR) in CACS for the two scans performed on the same patient 30 min apart was 51 (0, 100) (Table 3). Of the 22 patients that had a baseline score $>$ 30, the median absolute difference (IQR) and the median percent absolute difference (IQR) in CACS between the two scans were 71 (42, 115) and 10% (5%, 16%), respectively. Six (27%) of the 22 patients had an $>15\%$ change in CACS following repositioning. Four of these patients had an increase in CACS with position change [18% (95% CI: 5–40%)] and

Fig. 1. Patient breakdown by the calcification score for overall cohort as well as by CKD stage. For the comparisons between calcification score groups, $P = 0.82$.

Table 2. Reviewer agreement

Values are expressed as median [IQR].

Table 3. Inter-scan reproducibility

Values are expressed as median [IQR].

two had a decrease. Figure 2 shows that the difference was greatest in those with higher relative baseline CACS and Figure 3 shows that the percent difference was greatest in those with lower relative baseline CACS.

Change in calcification score at 1 year

Of the 33 patients in the analysis cohort, 22 patients underwent a 1-year follow-up scan. Two of these patients had undergone interim coronary artery bypass grafting and were unable to have CACS determined because of interference from coronary artery bypass graft clips. Reasons for not having a follow-up scan included patient moved (1), refused (8), and died (2). Patients who did not have a follow-up scan had higher baseline CAC [937 (IQR: 279, 2515) versus 236 (IQR: 0, 599)], lower HDL [42.5 mg/dL (SD = 7.7) versus 58 mg/dL $(SD = 23.2)$] and a higher percentage of history of CAD (59% versus 15%).

Seven of the patients who underwent a follow-up scan had a 0 CACS at baseline. None of these patients showed progressive calcification. Of the remaining 13 patients, 7 [54% (95% CI 25–81%)] showed progression of CAC and 2 (15%) had an apparent remission of CAC, defined as an increase and decrease, respectively, of at least 15% change from baseline. The median difference (IQR) in CACS for the entire cohort undergoing a follow-up scan was

Fig. 2. Change in CACS following repositioning (30-min scan) versus baseline CACS (in patients with >30 CACS at baseline). (--) indicates the median change in CACS following repositioning versus baseline CACS.

Fig. 3. Percent change following repositioning (30-min scan) versus baseline CACS (in patients with calcification scores >30 at baseline). (––) is drawn at ± 15 %, the level of change considered to represent true change from baseline.

16 (0, 200). Excluding the patients with a baseline CACS <30 from the analysis, the median difference (IQR) in CACS was $85 (-2, 218)$, and the median percent difference was 16% (−0.4%, 63%) (Table 4). Figure 4 compares the change in CACS scores at 30 min after baseline to change at 1 year.

Discussion

This study explores the reliability and variability of HRCTdetermined CACS in 33 patients with CKD. We demonstrate that a significant change in CACS can occur simply with patient repositioning and with the observer. The strengths of the study include a rigorous examination of the reliability of this test using measures of intra- and inter-

Table 4. Change in CACS at 1 year

	Change in CACS at 1 year	
All scans		
Number of scans	20	
Difference in CAC score (AU)	16 [0, 200]	
CAC score > 30		
Number of scans	13	
$>15\%$ change in CAC from baseline (%)	69	
Difference in CAC score (AU)	$85[-2, 218]$	
Percent difference (%)	$16[-0.4, 63]$	

Values are expressed as median [IQR].

reviewer reliability from multiple blind reviewers, as well as the test and retest within one time period. The observed

Fig. 4. Distribution of percentage change at 30 min after baseline and at 1 year (in patients with calcification scores $>$ 30 at baseline); $[1] = 95\%$ exact confidence interval. Note the overlapping confidence intervals for the percentage of patients that have >15% increase in CACS from baseline with repositioning and at 1 year.

test–retest variation in CACS questions the use of HRCTderived CACS as a surrogate outcome in clinical trials.

The Agatston score [11] is the most widely used and the best established measure of CAC. In the general population, the normal progression of the CACS per year is reported to be 14–27% [15]. The median inter-scan Agatston score variability has been reported to be in the range of 8–37% [14–17], which limits the detection of change within this range. Factors influencing inter-scan variability include partial volume effect, the use of the step function in the Agatston calculation, coronary artery motion, image noise, field inhomogeneity, lack of calibration and total amount of CAC [17]. The reliability of the CACS is not documented in patients with CKD. Regardless, progression of CAC, defined as a change in CACS of $>15\%$, has been used as an outcome measure for interventional trials in patients with CKD. We found significant variability when comparing two scans performed on the same patient 30 min apart, with 27% of scores showing an $>15%$ change from baseline. Of those with a $>15\%$ change from baseline, four patients had an increase in CACS, the degree of which would be consistent with 'progression' in interventional studies [12,13].

Methods to calculate change in calcification are challenging due to the fact that different absolute or percent changes have different implications at different ranges [12,13] (i.e. an increase in CACS from 50 to 100 is an absolute change of 50 but a percent change of 100%, whereas an increase from 450 to 900 is an absolute change of 450 but the same percent change). It is not surprising that we found that the error was greatest in patients with a baseline CACS in the high range when evaluated on the natural scale (figure 2) and in the mild–moderate range when evaluated on the percent scale (Figure 3). Notably, most patients with CKD have mild–moderate calcification. Use of percent change to calculate progression in this patient group, as occurred in the above-mentioned studies [12,13], risks overestimating progression.

The majority of studies of reliability in the general population have demonstrated inter- and intra-reviewer scoring result agreement of >90% [18,19]. We demonstrated similar reviewer agreement for the scans performed at baseline, but noted substantially lower inter-reviewer agreement for the 12-month scan. At 12 months, fewer scans were performed and they were reviewed over a more extended period, making it possible that reduced reviewer familiarity with scan reading may be responsible for this result. Thus, while our results suggest that reviewers can reliably score scans, they highlight the importance of reviewer experience.

We demonstrated progression of CAC at 1 year in 30% of the patients, consistent with previous reports [7,20] of accelerated progression in CKD patients. Of note, of the seven patients without CAC at baseline, none had evidence of calcification on the follow-up scan. Previous studies have reported a similar lack of calcification in those with no disease initially [8,12,21]. However, given our findings of inter-scan reproducibility and observer error, it is possible that a proportion of these patients 'progressed' as a consequence of error inherent in measurement process itself.

Our findings raise questions regarding the validity of the conclusions drawn in clinical practice and research studies [12,13] that are based on the current definition of CACS progression (>15% change from baseline). The CACS variation with scans taken on the same day suggests that it may be appropriate to modify the threshold for clinically significant differences in CACS. Specifically, a higher threshold to indicate change may be required, or perhaps what constitutes a significant change in calcification should depend on the baseline CACS. Alternatively, different methods for quantifying CAC [14], or newer, more precise scanning techniques [16], may help address the problem of variability. Regardless, measures of variability and precision of the scans should be incorporated into study design.

The primary limitation of this study is the small sample size. Nevertheless, our results are concordant with previous studies. Second, we did not assess CAC using the volume

score or calcium mass, despite suggestions that these methods may be superior to the Agatston score [14]. However, these methods have been less well studied and are less often reported, and patient management based on these measures is difficult because of paucity of representative data on CAC distribution. Third, we scanned patients with HRCT despite most available data in the literature on significant Agatston score changes being based on electron beam rather than HRCT. However, HRCT has been shown to have good correlations with electron beam CT in coronary artery calcium measurement [17,22], with some studies showing higher reproducibility with HRCT [17]. Fourth, only 22 of 33 patients received a repeat scan at 30 min, and only 20 of 33 patients had a valid repeat scan at 1 year. While there was no significant difference between patients who did and did not undergo the 30-min control scan, the patients who did not undergo a 12-month scan had risk factors for progression. Therefore, our results pertaining to progression are likely to be an underestimation of the effect and need to be interpreted with caution.

In summary, this small study describes a wellcharacterized cohort of patients with CKD, in whom 24% have no coronary calcification, 76% have abnormal CACS and 30% have progression of CACS over a 12-month period according to the currently accepted definition. However, it simultaneously raises questions regarding the utility of CACS as a surrogate outcome measure in clinical trials, given that some patients who had repeated scans on the same day demonstrated score variations of the same magnitude as is considered significant in long term studies $(i.e. >15\%$ change from baseline). At the least, this study demonstrates the importance of standardizing and understanding the characteristics of tests prior to using them in clinical trials.

Conflict of interest statement. None declared.

References

- 1. Parfrey PS, Foley RN. The clinical epidemiology of cardiac disease in chronic renal failure. *J Am Soc Nephrol* 1999; 10: 1606–1615
- 2. Stack AG, Bloembergen WE. A cross-sectional study of the prevalence and clinical correlates of congestive heart failure among incident US dialysis patients. *Am J Kidney Dis* 2001; 38: 992–1000
- 3. Ganesh SK, Stack AG, Levin NW *et al*. Association of elevated serum PO(4), Ca \times PO(4) product, and parathyroid hormone with cardiac mortality risk in chronic hemodialysis patients. *J Am Soc Nephrol* 2001; 12: 2131–2138
- 4. Block GA, Hulbert-Shearon TE, Levin NW *et al*. Association of serum phosphorus and calcium \times phosphate product with mortality risk in chronic hemodialysis patients: a national study. *Am J Kidney Dis* 1998; 31: 607–617
- 5. O'Rourke RA, Brundage BH, Froelicher VF *et al*. American College of Cardiology/American Heart Association Expert Consensus document on electron-beam computed tomography for the diagnosis and prognosis of coronary artery disease. *Circulation* 2000; 102: 126– 140
- 6. Kramer H, Toto R, Peshock R *et al*. Association between chronic kidney disease and coronary artery calcification: the Dallas heart study. *J Am Soc Nephrol* 2005; 16: 507–513
- 7. Tomiyama C, Higa A, Dalboni MA *et al*. The impact of traditional and non-traditional risk factors on coronary calcifiation in pre-dialysis patients. *Nephrol Dial Transplant* 2006; 21: 2464–2471
- 8. Russo D, Palmiero G, De Blasio A *et al*. Coronary artery calcification in patients with CRF not undergoing dialysis. *Am J Kidney Dis* 2004; 44: 1024–1030
- 9. Braun J, Oldendorf, M, Moshage, W *et al*. Electron beam computed tomography in the evaluation of cardiac calcifications in chronic dialysis patients. *Am J Kidney Dis* 1996; 27: 394–401
- 10. Levey AS, Bosch JP, Lewis JB *et al*. A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. *Ann Intern Med* 1999; 130: 461–470
- 11. Agatston AS, Janowitz Hildner WR FJ *et al*. Quantification of coronary artery calcium using ultrafast computed tomography. *J Am Coll Cardiol* 1990; 15: 827–832
- 12. Chertow GM, Burke, SK Raggi, P. Sevelamer attenuates the progression of coronary and aortic calcification in hemodialysis patients. *Kidney Int* 2002; 62: 245–252
- 13. Block GA, Speigal DM, Ehrlich J *et al*. Effects of sevelamer and calcium on coronary artery calcification in patients new to hemodialysis. *Kidney International* 2005; 68: 1815–1824
- 14. Callister TQ, Cooil B, Raya SP *et al*. Coronary artery disease: improved reproducibility of calcium scoring with an electron-beam CT volumetric method. *Radiology* 1998; 208: 807–814
- 15. Horiguchi J, Shen Y, Akiyama Y *et al*. Electron beam CT versus 16- MDCT on the variability of repeated coronary artery calcium measurements in a variable heart rate phantom. *Am J Roentgenol* 2005; 185: 995–1000
- 16. Kopp AF, Ohnesorge B, Becker C *et al*. Reproducibility and accuracy of coronary calcium measurements with multi-detector row versus electron-beam CT. *Radiology* 2002; 225: 113–119
- 17. Horiguchi J, Yamamoto H, Akiyama Y *et al*. Variability of repeated coronary artery calcium measurements by 16-MDCT with retrospective reconstruction. *Am J Roentgenol* 2005; 184:1917– 1923
- 18. Yamamoto H, Budoff MJ, Lu B *et al*. Reproducibility of three different scoring systems for measurement of coronary calcium. *Int J Cardiovasc Imaging* 2002; 18: 391–397
- 19. Shields JP, Miekle CH Jr, Watson P. Inter-rater reliability of electron beam computed tomography to detect coronary artery calcification. *Am J Card Imaging* 1996; 10: 97–100
- 20. Bursztyn M, Motro M, Grossman E *et al*. Accelerated coronary artery calcification mildly reduced renal function of high risk hypertensives: a 3-year prospective observation. *J Hypertens* 2003; 21: 1953–1959
- 21. Goodman W, Goldin, J, Kuizon, B *et al*. Coronary-artery calcification in young adults with end-stage renal disease who are undergoing dialysis. *N Engl J Med* 2000; 342: 1478–1483
- 22. Daniell A, Wong N, Friedman JD *et al*. Concordance of coronary artery calcium estimates between MDCT and electron beam tomography. *Am J Roentgenol* 2005; 185: 1542–1545

Received for publication: 23.12.07 Accepted in revised form: 4.4.08