
Interobserver Agreement on Captopril Renography for Assessing Renal Vascular Disease

Pieta Krijnen, MSc¹; Hong-Yoe Oei, MD, PhD²; Roel A.M.J. Claessens, MD, PhD³; Jan C. Roos, MD, PhD⁴; Brigit C. van Jaarsveld, MD, PhD⁵; and J. Dik F. Habbema, PhD¹

¹Center for Clinical Decision Sciences, Department of Public Health, Erasmus University Rotterdam, Rotterdam, The Netherlands;

²Department of Nuclear Medicine, University Hospital Rotterdam-Dijkzigt, Rotterdam, The Netherlands; ³Department of Nuclear

Medicine, Bosch Medicentrum, 's Hertogenbosch, The Netherlands; ⁴Department of Nuclear Medicine, University Hospital Vrije

Universiteit, Amsterdam, The Netherlands; and ⁵Dianet Dialysis Centers, Utrecht, The Netherlands

Captopril-stimulated renography is widely used to screen selected groups of hypertensive patients for renal vascular disease. Evaluation of the test is a complex task. Lack of interobserver agreement on the assessment and interpretation of renographic parameters may contribute to differences in sensitivity and specificity between studies. **Methods:** Three experienced nuclear medicine physicians evaluated 658 renograms of 503 hypertensive patients suspected of having renal vascular disease from a large Dutch multicenter study (the Dutch Renal Artery Stenosis Intervention Cooperative [DRASTIC] study). Interobserver agreement on several renographic parameters was assessed by the κ statistic and the intraclass correlation coefficient (ICC). **Results:** The interobserver agreement on the time to excretion was high: The pooled ICC was 0.90. The pooled κ was ≥ 0.65 for the pattern of the time-activity curves, the visual aspect of the scintigraphic images (visible uptake and kidney size), and the judgment on the presence of renal artery stenosis. However, the interobserver agreement on cortical retention and pelvic retention by visual inspection of the images was rather low (pooled $\kappa = 0.46$ and 0.52 , respectively). Pelvic retention was found to complicate the interpretation of renography. **Conclusion:** Interobserver agreement on most of the renographic parameters was satisfactory, but the assessment of cortical retention was more difficult, in particular, in the presence of pelvic retention. Captopril renography should be interpreted with caution if pelvic retention is suspected. Interobserver variability offers one of several explanations for the differences in diagnostic test performance that are found between studies.

Key Words: renography; interobserver agreement; renovascular hypertension

J Nucl Med 2002; 43:330–337

Captopril-stimulated renography is a noninvasive test that is widely used to screen selected groups of hypertensive patients for the presence of renal vascular disease. In patients with renovascular hypertension, captopril induces changes in the scintigraphic images of the kidney distal to the stenosis by revealing decreased uptake or delayed excretion with cortical retention (or both). Accordingly, the time-activity curves may reveal these alterations. Evaluation of scintigraphic images and time-activity curves is encouraged in the investigation of renal vascular disease (1–3). Patients with such captopril-induced changes on the renogram are generally expected to benefit from intervention with balloon angioplasty or with stent insertion (4–6).

Interpretation of captopril renography is not a straightforward task. The nuclear medicine physician must assess several renographic parameters and subsequently integrate this information to form a judgment on the presence of renal vascular disease. Efforts have been made to standardize the test (1,7–9). These guidelines focus mainly on the procedure and not on interpretation of the results. Moreover, diagnostic criteria are not uniform, and different renographic parameters are considered. The diagnostic performance of captopril renography has been variously described with sensitivity ranging between 70% and 100% and specificity ranging between 60% and 100% (6,10,11). A lack of interobserver agreement on interpretation of the test results may have contributed to these differences. Despite the vast literature on captopril renography for diagnosing renal vascular disease, the interobserver variability has not yet been described.

In this study, 3 experienced nuclear medicine physicians, working in different university hospitals, evaluated 658 renograms of 503 patients suspected of having renal vascular disease. We analyzed the interobserver agreement on the assessment of renographic parameters and the agreement on the judgment on the presence of hemodynamically significant renal artery stenosis.

Received Aug. 7, 2001; revision accepted Dec. 3, 2001.
For correspondence contact: Pieta Krijnen, MSc, Center for Clinical Decision Sciences, Department of Public Health, Erasmus University Rotterdam, P.O. Box 1738, Rotterdam, 3000 DR, The Netherlands.
E-mail: krijnen@mgz.fgg.eur.nl

MATERIALS AND METHODS

Study Design

The study was part of the Dutch Renal Artery Stenosis Intervention Cooperative (DRASTIC) study. The aim of this multicenter study was to optimize the diagnosis and treatment of renal artery stenosis (12,13). The DRASTIC study included 1,205 hypertensive patients, 18–75 y old, who had been referred for unsatisfactory control of blood pressure or an adverse drug effect during the course of antihypertensive treatment or for analysis of possible secondary hypertension. Exclusion criteria were suspected secondary hypertension other than renal vascular disease, unstable coronary artery disease, heart failure, renal failure (serum creatinine, $\geq 200 \mu\text{mol/L}$ [2.26 mg/dL]), and inadequate contraception. Patients with drug-resistant hypertension (diastolic blood pressure, $\geq 95 \text{ mm Hg}$ on 2 drugs during 3 visits) ($n = 455$) or with a rise in serum creatinine concentration after angiotensin-converting enzyme (ACE) inhibitor therapy ($n = 43$) as well as patients in whom renal artery stenosis had been diagnosed before their referral to the participating center ($n = 72$) underwent diagnostic workup for renal artery stenosis. Patients with atherosclerotic renal artery stenosis, defined as $\geq 50\%$ reduction of lumen diameter according to renal digital subtraction angiography (gold standard test), were randomly assigned to either the balloon angioplasty ($n = 56$) or the medical treatment ($n = 50$) group. Captopril renography was performed and evaluated by the local nuclear medicine physicians in 22 participating hospitals. In the diagnostic workup, the sensitivity and specificity for finding stenosis according to the local nuclear medicine physician were 72% and 90%, respectively (11). Furthermore, renography was performed to evaluate treatment after 3 and 12 mo of follow-up.

Renographic Protocol

The protocol for conducting the renographic procedures reflected the guidelines of the consensus report on ACE inhibitor renography (7). In patients who were receiving long-term ACE inhibitor treatment, the ACE inhibitor was withheld for at least 24 h before renography was performed. According to the protocol, an oral dose of 50 mg captopril was given 1 h before the examination in 95% of the procedures to induce asymmetry in uptake and intrarenal transit between the kidneys in the case of renal vascular disease. In the remaining 5% of the procedures, the physician reduced the dose of captopril to 25 mg to prevent hypotension. To ensure adequate absorption of captopril, patients were required to fast during the 4 h preceding renography. Sufficient hydration was guaranteed by oral administration of 0.5 L of tap water. Blood pressure was measured with an automatic device before administration of captopril and every 5–10 min for 2 h after administration of captopril. Renography was performed with the patient in supine position and the detector placed posteriorly. After intravenous administration of 75–100 MBq $^{99\text{m}}\text{Tc}$ -mercaptoacetyl-triglycine, data were collected in 10-s frames during a 20-min period, and sequential analog images were obtained every minute. The time–activity curves were generated using regions of interest over the whole kidney (1).

Study

In this study on interobserver agreement, 658 renograms of 503 patients with 2 native kidneys were reevaluated by 3 experienced nuclear medicine physicians (referred to as physicians A, B, and C) who were working in different university hospitals at the time. Of these renograms, 487 were obtained during the diagnostic

workup of patients with and without renal artery stenosis. The remaining renograms were obtained during follow-up of patients with stenosis: 82 renograms after 3 mo of follow-up and 89 renograms after 12 mo of follow-up.

Renographic Evaluation

The renograms were evaluated independently, and the physicians were unaware of patient characteristics and hospital source. The 3 physicians had no additional clinical information, such as the blood pressure response to captopril and the diuresis during the procedure. Before evaluation, the physicians discussed which renographic parameters of the scintigraphic images and time–activity curves would be assessed and how these features would be scored.

The following parameters were scored from the scintigraphic images by each individual observer, separately for the left and right kidneys: visible uptake (scored as present or absent); time to excretion (scored as number of minutes until radioactivity appeared in the renal pelvis, determined by visual evaluation of the 1-min sequential images, if available; if the excretory phase started only after 20 min, no excretory phase was registered); and kidney size (scored as normal or small). Cortical retention and pelvic retention (scored as present or absent) were determined by visual inspection. The presence of pelvic retention was assessed because this was considered to complicate the renographic evaluation of the images and the time–activity curves of the whole kidneys (1,14). The pattern of the time–activity curves was scored in 6 ordered categories as proposed by Fommei et al. (10) (0 = normal, 1 = minor abnormalities, 2 = marked delayed excretion rate with preserved washout phase, 3 = delayed excretion rate without washout phase [accumulation curve], 4 = renal failure pattern with measurable kidney uptake, and 5 = renal failure pattern without measurable kidney uptake [blood background-type curve]). Interobserver agreement was not applicable for the time to peak activity (T_{max}) and the relative (individual kidney) uptake because these diagnostic criteria were calculated by the computer.

Finally, the judgment on the presence or absence of renal artery stenosis was assessed for each kidney. No specific diagnostic criteria were defined to reflect the clinical practice. The judgment on the presence of stenosis was scored as 1 of 5 ordered categories (1 = certainly stenosis, 2 = probably stenosis, 3 = indeterminate, 4 = probably no stenosis, and 5 = certainly no stenosis; in the case of a blood background-type curve, the diagnosis was scored as indeterminate).

Interobserver Agreement

We used the κ statistic to assess interobserver agreement on the renographic parameters that were measured on a nominal scale. κ reflects the proportion of the maximally achievable agreement that is realized on top of the agreement that is expected by chance (15–17). κ values usually range from 0 (indicating chance agreement only) to 1 (indicating perfect agreement). The only meaningful interpretation of negative values of κ is that the level of agreement is what would be expected by chance alone (18). In general, κ values of < 0.40 are considered as low and values of > 0.80 are considered as high (16,17,19). Because the value of κ decreases if the number of ordinal categories is increased, we calculated weighted κ values for the pattern of the time–activity curves and the judgment on the presence of stenosis to adjust for the seriousness of different levels of disagreement (20–22). Linear weights were used: $w(ij) = 1 - |i - j|/(c - 1)$, where i and j are the sequence numbers of the categories, and c is the number of

categories. Interpretation of weighted κ is like that of unweighted κ (16).

The interobserver agreement on the time to excretion, which was measured on an interval scale, was expressed as the intraclass correlation coefficient (ICC). The ICC takes into account systematic differences between observers and ranges from -1 (perfect disagreement) to 1 (perfect agreement), with 0 indicating only random concordance (23,24). Although there are no universal standards, values of ICC of <0.40 are considered as low and values of >0.75 are considered as high (25).

Interobserver agreement on renographic parameters was calculated by kidney and on the judgment on the presence of stenosis by kidney as well as by patient. Interobserver agreement was assessed for each pair of observers. A pooled estimate was also calculated on the basis of the mean observed agreement and the mean amount of agreement expected under the null model of independence. A 95% confidence interval (CI) was calculated for each estimate. Estimates of the ICC were calculated with SPSS software (release 9.0.0; SPSS, Chicago, IL) and estimates of κ were calculated with Agree Statistical Software (version 7.001; ProGAMMA, Groningen, The Netherlands).

Finally, the probability that a physician judged stenosis to be absent given the fact that another did so, corrected for chance agreement, was calculated using the average conditional probability of the absence of stenosis and the average expected probability of stenosis. Similar probabilities were calculated for the judgment on the presence of unilateral stenosis and for the judgment on the presence of bilateral stenosis. These probabilities can be interpreted as a κ -per-outcome category.

RESULTS

Patients

All patients whose renograms were evaluated had drug-resistant hypertension. Their diastolic blood pressure was 105 ± 9 mm Hg (mean \pm SD) despite the use of 2 ± 1 antihypertensive drugs. At study entry, the renal function was normal or mildly impaired: The patients had a serum creatinine concentration of 95 ± 27 $\mu\text{mol/L}$ and their creatinine clearance was 85 ± 33 mL/min (Table 1). In 5 patients the serum creatinine concentration had increased to >150 $\mu\text{mol/L}$ during follow-up.

TABLE 1
Characteristics at Entry of 503 Patients Evaluated
for Renal Artery Stenosis

Characteristic	Percentage of patients or mean \pm SD
Age (y)	52 ± 13
Male (%)	57
Systolic blood pressure (mm Hg)	171 ± 23
Diastolic blood pressure (mm Hg)	105 ± 9
No. of antihypertensive drugs	2 ± 1
Serum creatinine ($\mu\text{mol/L}$)	95 ± 27
Creatinine clearance (mL/min)	85 ± 33
Referred by general practitioner (%)	51
Stenosis $\geq 50\%$ on angiography (%)	30

Scintigraphic Images

The 3 nuclear medicine physicians did not note any uptake on the scintigraphic images of 1%–3% of the kidneys (Table 2). Physician C reported the absence of uptake twice as often as physicians A and B. The pooled κ -value for visual uptake was 0.65 (95% CI, 0.51–0.80). A small kidney size was scored more frequently by physician A than by the other physicians (25% vs. 18% and 17%). The pooled κ was 0.70 (95% CI, 0.66–0.74). Because 1-min images were not obtained routinely in every hospital, the beginning of the excretory phase was assessed for approximately half of the renograms. The beginning of the excretory phase was estimated to start, on average, after 4.29–4.43 min. The pooled ICC was 0.90 (95% CI, 0.89–0.91). Cortical retention was reported in 2–3 times as many kidneys by physician A as by the other physicians. The pooled κ was 0.46 (95% CI, 0.42–0.51). Pelvic retention was reported least by physician B (12% vs. 21% and 18%). The pooled κ for pelvic retention was 0.52 (95% CI, 0.47–0.56).

Time-Activity Curves

Systematic differences occurred between observers in assigning a pattern to the time-activity curves (Table 2; Fig. 1). Physician C reported more abnormal time-activity curves than physicians A and B. Furthermore, physician A reported more abnormal curves than physician B. The pooled value for the weighted κ was 0.65 (95% CI, 0.62–0.68).

Judgment on Presence of Stenosis by Kidney

The pooled value of the weighted κ for the judgment on the presence of stenosis for separate kidneys, as measured on a 5-point scale, was 0.16 (95% CI, 0.13–0.18) (Table 2). Physician B was more outspoken in assigning scores than physicians A and C: Physician B was certain of the presence of stenosis in 4% of the kidneys compared with 2% and $<1\%$ (physicians A and C, respectively) and was certain of the absence of stenosis in 59% of the kidneys compared with 6% and 18% (physicians A and C, respectively) (Fig. 2). When the judgment on the presence of stenosis was dichotomized into certainly or probably stenosis or indeterminate versus certainly or probably no stenosis, an indication for stenosis was found in 14%–22% of the kidneys. The pooled κ for the dichotomized judgment was better than that on the 5-point scale: 0.66 (95% CI, 0.62–0.70) versus 0.16 (95% CI, 0.13–0.18).

κ was calculated separately for those kidneys on which all 3 physicians agreed that pelvic retention had or had not occurred. For kidneys showing pelvic retention ($n = 90$), κ for the dichotomized judgment on the presence of stenosis was significantly lower than that for kidneys without pelvic retention ($n = 909$): κ ranged between -0.07 and 0.12 for kidneys with pelvic retention (pooled estimate, 0.06; 95% CI, -0.04 to 0.15) and between 0.69 and 0.77 for kidneys without pelvic retention (pooled estimate, 0.73; 95% CI, 0.68–0.78).

TABLE 2
Interobserver Agreement on Renographic Parameters

Renographic parameter (no. of kidneys studied)	Physician	No. (%) with feature or mean \pm SD	Physicians	Agreement*
Scintigraphic images				
No visible uptake ($n = 1,306$)	A	14 (1.1)	A and B	0.81 (0.66–0.96)
	B	18 (1.4)	B and C	0.59 (0.43–0.76)
	C	32 (2.5)	A and C	0.60 (0.44–0.77)
			Pooled	0.65 (0.51–0.80)
Small kidney ($n = 1,271$)	A	323 (25)	A and B	0.69 (0.64–0.74)
	B	229 (18)	B and C	0.79 (0.74–0.83)
	C	210 (17)	A and C	0.64 (0.59–0.69)
			Pooled	0.70 (0.66–0.74)
Time to excretion (min) ($n = 708$)	A	4.29 \pm 1.47	A and B	0.88 (0.86–0.90)
	B	4.34 \pm 1.65	B and C	0.95 (0.94–0.96)
	C	4.43 \pm 1.67	A and C	0.88 (0.86–0.89)
			Pooled	0.90 (0.89–0.91)
Cortical retention ($n = 1,238$)	A	350 (28)	A and B	0.37 (0.31–0.42)
	B	119 (10)	B and C	0.63 (0.56–0.70)
	C	165 (13)	A and C	0.46 (0.41–0.52)
			Pooled	0.46 (0.42–0.51)
Pelvic retention ($n = 1,256$)	A	257 (21)	A and B	0.48 (0.41–0.54)
	B	147 (12)	B and C	0.46 (0.40–0.53)
	C	231 (18)	A and C	0.59 (0.54–0.65)
			Pooled	0.52 (0.47–0.56)
Time-activity curves:				
Pattern of curve ($n = 1,274$)	A	†	A and B	0.74 (0.70–0.77)
	B	†	B and C	0.60 (0.57–0.64)
	C	†	A and C	0.63 (0.60–0.67)
			Pooled	0.65 (0.62–0.68)
Conclusion on presence of stenosis				
On original 5-point scale‡ ($n = 1,316$)	A	§	A and B	0.05 (0.01–0.09)
	B	§	B and C	0.14 (0.10–0.17)
	C	§	A and C	0.32 (0.27–0.37)
			Pooled	0.16 (0.13–0.18)
Collapsed into 2-point scale¶ ($n = 1,316$)	A	293 (22)	A and B	0.62 (0.57–0.68)
	B	184 (14)	B and C	0.66 (0.60–0.71)
	C	256 (20)	A and C	0.69 (0.64–0.74)
			Pooled	0.66 (0.62–0.70)

* κ with 95% CI, except for time to excretion, where ICC with 95% CI is shown.

†See Figure 1.

‡1 = certainly stenosis; 2 = probably stenosis; 3 = indeterminate; 4 = probably no stenosis; 5 = certainly no stenosis.

§See Figure 2.

¶Indication for stenosis = certainly or probably stenosis or indeterminate; no indication for stenosis = certainly or probably no stenosis.

Judgment on Presence of Stenosis by Patient

The 3 physicians found an indication for stenosis (certainly or probably stenosis or indeterminate) on 20%–28% of the renograms. The pooled κ was 0.70 (95% CI, 0.64–0.76). Furthermore, we studied the agreement on whether there was no indication for stenosis or was an indication for unilateral stenosis or an indication for bilateral stenosis (Fig. 3). An indication of bilateral stenosis was judged variously: Physician B suspected bilateral stenosis in 4% of the patients, whereas physicians A and C suspected

bilateral stenosis to be present more frequently (in 12% and 11%, respectively). When 1 of the 3 physicians judged that stenosis was absent, the probability that a second physician concluded the same was, on average, 0.70 (95% CI, 0.61–0.79). When 1 of the 3 physicians judged that unilateral stenosis was present, the probability that a second physician reached the same conclusion was, on average, 0.65 (95% CI, 0.61–0.70). For the presence of bilateral stenosis, this probability was, on average, 0.48 (95% CI, 0.43–0.52).

Physician A		0	1	2	3	4	5	Total
Physician B	Pattern							
	0	735	141	1		2		879
	1	14	156	45		4		219
	2	4	24	60	5	3		96
	3			12	42			54
	4	3	4	3	3	4	4	21
	5					2	3	5
Total	756	325	121	50	15	7	1274	

Physician B		0	1	2	3	4	5	Total
Physician C	Pattern							
	0	508	7					515
	1	351	116	3		1		471
	2	14	91	69	3	4		181
	3		1	20	51	1		73
	4	6	4	4		12		26
	5					3	5	8
Total	879	219	96	54	21	5	1274	

Physician A		0	1	2	3	4	5	Total
Physician C	Pattern							
	0	496	19					515
	1	250	205	16				471
	2	8	93	72	4	4		181
	3			28	42	3		73
	4	2	8	5	2	6	3	26
	5				2	2	4	8
Total	756	325	121	50	15	7	1274	

* 0=normal, 1=minor abnormalities, 2=marked delayed excretion rate with preserved washout phase, 3=delayed excretion rate without washout phase (accumulation curve), 4=renal failure pattern with measurable kidney uptake, 5=renal failure pattern without measurable kidney uptake (blood background-type curve)

FIGURE 1. Interobserver agreement on pattern of time-activity curves for 1,274 kidneys according to 3 nuclear medicine physicians.

DISCUSSION

In this study, the interobserver agreement on captopril renography was studied on 658 renograms of patients with drug-resistant hypertension and a normal or mildly impaired renal function. Three experienced nuclear medicine physicians assessed renographic parameters that have been recommended for evaluation (1,2,7) and judged whether hemodynamically significant renal artery stenosis was present or absent. For most of these parameters and for the judgment on presence of stenosis, the interobserver agreement was satisfactory. The agreement on cortical retention was relatively low, however.

Except for the time to excretion, the interobserver agreement was assessed by the κ statistic. Although κ is most commonly used to measure interobserver agreement in categorical data, one has to bear in mind, however, that the interpretation of κ is complicated by some of its properties (17,18,26). First, the value of κ strongly depends on the

underlying prevalence of the parameter under study. For instance, a high value of κ for agreement on the absence of visible uptake is harder to achieve than for agreement on small kidney size because the latter is much more common. Second, although κ does not identify systematic differences between observers (bias), κ will be lower if such bias is present. This is also the case for the ICC, which was used to assess interobserver agreement in continuous data. Therefore, it should be noted that systematic differences between the observers in the assessment of several parameters were found—for instance, for the judgment on the presence of stenosis (Fig. 2). Third, the way one values discrepancies between categories and consequently chooses the weights for the calculation of the weighted κ is arbitrary. For instance, by choosing linear weights in the calculation of κ for the time-activity curves, we assumed that disagreement between normal curves and curves with minor abnormalities (curve types 0 and 1) is as serious as disagreement between

Physician A		1	2	3	4	5	Total
Physician B	Conclusion*						
	1	13	40	5			58
	2	2	22	14	6	1	45
	3	7	31	30	11	2	81
	4	1	47	47	264	1	360
	5		5	29	664	74	772
Total	23	145	125	945	78	1316	

Physician B		1	2	3	4	5	Total
Physician C	Conclusion*						
	1	1		1			2
	2	35	11	25	16		87
	3	21	23	40	67	16	167
	4	1	10	13	262	543	829
	5		1	2	15	213	231
Total	58	45	81	360	772	1316	

Physician A		1	2	3	4	5	Total
Physician C	Conclusion*						
	1	2					2
	2	16	65	4	2		87
	3	4	54	62	44	3	167
	4	1	24	55	707	42	829
	5		2	4	192	33	231
Total	23	145	125	945	78	1316	

* 1=certainly stenosis, 2=probably stenosis, 3=indeterminate, 4=probably no stenosis, 5=certainly no stenosis

FIGURE 2. Interobserver agreement on judgment on presence of renal artery stenosis for 1,316 kidneys according to 3 nuclear medicine physicians.

renal failure patterns with and without measurable kidney uptake (curve types 4 and 5).

The pattern of the time-activity curves, which is considered to be an important diagnostic parameter (2,14), was scored in 6 ordered categories (10). The weighted κ value for the pattern of the time-activity curves was moderately high, especially when one considers that the distinction between some of these types of curves is difficult to make. The interobserver agreement on visible uptake and on kidney size was also satisfactory but could have been affected negatively by the low prevalence of these features (18,26). The interobserver agreement on time to excretion as assessed from the scintigraphic images was high. Yet, the relative (individual kidney) uptake and the T_{max} are the most reliable parameters in terms of interobserver agreement because the computer calculates them.

With ^{99m}Tc -mercaptoacetyltryglycine, which is almost completely cleared by tubular secretion, renovascular hypertension can usually be detected by cortical retention after ACE inhibition (7). Delayed excretion can also be caused by pelvic stasis, however. In kidneys without a dilated renal pelvis, pelvic retention will be observed because of low diuresis. The patients in this study drank 0.5 L of tap water

30–60 min before the renography. Perhaps a more abundant diuresis could be achieved by giving 10 mL/kg of body weight. Another cause of low diuresis is the fact that some of the patients were on diuretics. These patients may produce less urine during the renography (9). The identification of cortical retention is difficult in the presence of pelvic retention (1,14). The complicating role of pelvic retention in the evaluation of captopril renography was evident in our study. For cortical and pelvic retention, the interobserver agreement on the assessment of the presence or absence of these phenomena was not satisfactory. Probably, this can be improved by the assessment of the time-activity curves of the renal cortex.

Which renographic parameters should be used then as diagnostic criteria in the evaluation of renal vascular disease? The diagnostic performance and the interobserver variability should be included in this consideration. When ranked according to the sum of sensitivity and specificity in a by-patient analysis, the order of the renographic parameters was virtually the same for the 3 nuclear medicine physicians (data not shown). However, one must bear in mind that by this way of ranking the sensitivity and the specificity are valued equally. The parameter with the best

		Physician A			
		Conclusion	No RAS	URAS	BIRAS
Physician B	No RAS	387	26	27	440
	URAS	5	62	21	88
	BIRAS	4	1	18	23
	Total	396	89	66	551
	Physician B				
		Physician B			
		Conclusion	No RAS	URAS	BIRAS
Physician C	No RAS	401	13	4	418
	URAS	14	61		75
	BIRAS	25	14	19	58
	Total	440	88	23	551
	Physician A				
		Physician A			
		Conclusion	No RAS	URAS	BIRAS
Physician C	No RAS	375	28	15	418
	URAS	9	55	11	75
	BIRAS	12	6	40	58
	Total	396	89	66	551

FIGURE 3. Interobserver agreement on absence of renal artery stenosis (No RAS) or presence of unilateral (URAS) or bilateral (BIRAS) stenosis according to 3 nuclear medicine physicians on basis of conclusion per kidney.

diagnostic performance was asymmetry in renal uptake. The fact that the individual kidney uptake is measured objectively adds to its usefulness as a diagnostic criterion. Time to excretion as assessed from the scintigraphic images, an abnormal pattern of the time-activity curves, and cortical retention ranked somewhat lower in terms of diagnostic performance. On the basis of the interobserver variability of these parameters, the first 2 are also important diagnostic criteria but the last should be given less weight. The lowest diagnostic performance was found for the visual assessment of the kidneys on the scintigraphic images (i.e., no visible uptake or asymmetry in kidney size) and T_{max} . Diagnostic information is lost if one focuses on just 1 or 2 parameters when evaluating the test results. To maximize the diagnostic value of the test, all parameters might be brought together in multivariate models, one predicting the outcome of angiography and one predicting the response to treatment as primary outcome measures for the value of renography. These models may be used then to support decision making by nuclear medicine physicians. We intend to report on the usefulness of such models in the future.

The 3 evaluating physicians judged the presence of stenosis on a 5-point scale, which was collapsed into suspect or indeterminate versus not suspect to reflect which patients would normally be referred for further diagnostic workup. The interobserver agreement on the presence of stenosis was moderate. When 1 physician judged stenosis to be absent, the probability that a second physician concluded the opposite was 30%. It would seem that the interobserver agreement found in this study represents the maximum achievable because the evaluating physicians in this study

were well trained and experienced and had deliberated their way of scoring beforehand. On the other hand, the renograms were not always obtained according to the protocol (1-min images were not always acquired) and were not self-managed by the evaluating physicians. Also, to reflect the common clinical practice, diagnostic criteria for identifying stenosis were not specified before evaluation. Thus, the interobserver agreement found in this study could possibly be improved by performing the procedure and evaluation in a uniform manner.

CONCLUSION

The interobserver agreement on most renographic parameters was satisfactory. Important parameters for establishing the diagnosis of stenosis with high interobserver agreement were the relative (individual kidney) uptake, the pattern of the time-activity curves, and the time to excretion. The assessment of cortical retention by visual inspection of the images was more difficult—in particular, in the presence of pelvic retention—and should be given less weight in the evaluation. Captopril renography should be interpreted with caution if pelvic retention is present. Besides differences in patient selection, study design, and diagnostic criteria, interobserver variability offers an explanation for differences in diagnostic performance of captopril renography between studies.

ACKNOWLEDGMENT

This study was supported by grant OG92-031 from the Dutch Health Insurance Executive Board (Ziekenfondsraad).

REFERENCES

1. Oei HY. Captopril renography: early observations and diagnostic criteria. *Am J Hypertens*. 1991;4:678S–684S.
2. Nally JV Jr, Chen C, Fine E, et al. Diagnostic criteria of renovascular hypertension with captopril renography: a consensus statement. *Am J Hypertens*. 1991;4:749S–752S.
3. Nally JV Jr, Black HR. State-of-the-art review: captopril renography—pathophysiological considerations and clinical observations. *Semin Nucl Med*. 1992;22:85–97.
4. Oei HY, Geyskes GG, Mees EJ, Puylaert CB. The significance of captopril renography in renovascular hypertension. *Contrib Nephrol*. 1987;56:95–103.
5. Geyskes GG, Oei HY, Puylaert CB, Mees EJ. Renovascular hypertension identified by captopril-induced changes in the renogram. *Hypertension*. 1987;9:451–458.
6. Taylor A. Functional testing: ACEI renography. *Semin Nephrol*. 2000;20:437–444.
7. Taylor A, Nally J, Aurell M, et al. Consensus report on ACE inhibitor renography for detecting renovascular hypertension. *J Nucl Med*. 1996;37:1876–1882.
8. Taylor AT Jr, Fletcher JW, Nally JV Jr, et al. Procedure guideline for diagnosis of renovascular hypertension. *J Nucl Med*. 1998;39:1297–1302.
9. Oei HY. Dynamic and static renal imaging. In: Murray IPC, Ell PJ, eds. *Nuclear Medicine in Clinical Diagnosis and Treatment*. Edinburgh, U.K.: Churchill Livingstone; 1998:229–244.
10. Fommei E, Ghione S, Hilson AJ, et al. Captopril radionuclide test in renovascular hypertension: a European multicentre study. *Eur J Nucl Med*. 1993;20:617–623.
11. van Jaarsveld BC, Krijnen P. Prospective studies of diagnosis and intervention: the Dutch experience. *Semin Nephrol*. 2000;20:463–473.
12. van Jaarsveld B, Krijnen P, Bartelink A, et al. The Dutch Renal Artery Stenosis Intervention Cooperative (DRASTIC) Study: rationale, design and inclusion data. *J Hypertens Suppl*. 1998;16:S21–S27.
13. van Jaarsveld BC, Krijnen P, Pieterman H, et al. The effect of balloon angioplasty on hypertension in atherosclerotic renal-artery stenosis: Dutch Renal Artery Stenosis Intervention Cooperative Study Group. *N Engl J Med*. 2000;342:1007–1014.
14. Fommei E, Mezzasalma L, Ghione S, et al. European Captopril Radionuclide Test Multicenter Study: preliminary results—inspective renographic analysis. *Am J Hypertens*. 1991;4:690S–697S.
15. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20:37–46.
16. Fleiss JL. *Statistical Methods for Rates and Proportions*. New York, NY: Wiley; 1981:212–236.
17. Altman DG. *Practical Statistics for Medical Research*. London, U.K.: Chapman and Hall; 1991:403–409.
18. Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. *BMJ*. 1992;304:1491–1494.
19. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.
20. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull*. 1968;70:213–220.
21. Koran LM. The reliability of clinical methods, data and judgments: first of two parts. *N Engl J Med*. 1975;293:642–646.
22. Koran LM. The reliability of clinical methods, data and judgments: second of two parts. *N Engl J Med*. 1975;293:695–701.
23. Bartko JJ. On various intraclass correlation reliability coefficients. *Psychol Bull*. 1976;83:762–765.
24. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86:420–428.
25. Fleiss JL. *The Design and Analysis of Clinical Experiments*. New York, NY: Wiley; 1986:1–32.
26. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. the problems of two paradoxes. *J Clin Epidemiol*. 1990;43:543–549.

