Interpretation of Captopril Renography by Nuclear Medicine Physicians

Gerrit Schreij, Marinus J. van Kroonenburgh, Guido K. Heidendal, Harry A. van der Pol and Peter W. de Leeuw

Departments of Internal Medicine and Nuclear Medicine, University Hospital, Maastricht, The Netherlands; and Department of Nuclear Medicine, Maasland Hospital, Sittard, The Netherlands

This study was designed to assess intra- and interobserver variability and diagnostic accuracy of nuclear medicine physicians in their evaluation of baseline and captopril renograms. **Methods:** The diagnostic performance of three experienced nuclear medicine physicians according to their interpretation of baseline, captopril and paired renograms was assessed. To this end, the readers evaluated the renograms of 28 hypertensive patients in whom a diagnosis of renovascular hypertension was suspected on the basis of clinical clues. All patients also underwent angiography. The readers were unaware of the angiographic diagnosis. **Results:** Thirteen of 28 patients proved to have renal artery stenosis (8 unilateral, 5 bilateral) on renal angiography. The concordance in the renographic diagnoses between the three readers was reasonably good, with an intraobserver agreement and kappa (observed agreement corrected for chance) ranging from 64% to 89% and from 0.52 to 0.75, respectively, and an interobserver agreement and kappa ranging from 68% to 86% and from 0.61 to 0.82. The sensitivity of their interpretation of paired baseline plus captopril renograms in relation to the angiographic diagnosis is poor and below 50%. The post-test probability of RAS in case of a negative renographic study was found to be rather similar to the pre-test probability (prevalence) of 46%. Blinding readers to which renogram was obtained after captopril imaging increased their accuracy. **Conclusion:** The intra- and interobserver agreement between experienced nuclear medicine physicians who evaluate renograms was found to be reasonably good. Blinding readers as to which renogram is the pre- and post-captopril image seems to enhance their diagnostic accuracy in instances of positive scans.

Key Words: renography; renal artery stenosis; intraobserver agreement; inter-observer agreement; captopril

J Nucl Med 1995; 36:2192-2195

Physicians often have to rely on the results of diagnostic tests that indicate the presence or absence of disease or abnormal function. The interpretation of such tests, however, may show considerable bias and, hence, different investigators may not agree on their final diagnosis.

A common strategy to establish the reliability of a diagnostic test is to measure intra- and interobserver variability by asking one or more observers to evaluate the same test at two separate occasions. With this approach, the observed agreement for two cardiologists who examine the same electrocardiogram of different patients may be 57% with a kappa value (observed agreement corrected for chance) of 0.3 (*1*), while a 97% agreement with a kappa value of 0.67 for two radiologists examining the same set of mammograms may be found (*2*). To determine the level of intra- and interobserver agreement with respect to the renographic diagnosis of renal artery stenosis (RAS) in patients in whom there is a high clinical suspicion for this abnormality, we evaluated the visual interpretations of baseline and captopril renograms by three experienced nuclear medicine physicians.

In addition, we studied the accuracy of their renographic diagnosis in relation to the results of renal angiography, which was the gold standard (*3*).

METHODS

Patients

Twenty-eight consecutive hypertensive patients in whom renovascular hypertension was suspected on clinical grounds (*4*) underwent ^{99m}Tc-MAG-3 renography at baseline and 2 hr after an oral dose of 25 mg captopril. Before renography, all patients were given 300 ml of fluids to guarantee an urine output of at least 1 cc/min during the investigations. Subsequently, renal angiography was performed in all patients, irrespective of the renographic diagnosis.

Image Interpretation Protocol

The baseline and captopril renograms of all 28 patients were evaluated during two rounds by three independent nuclear medicine physicians (readers). They were unaware of the final angiographic diagnosis.

During the first round, the readers evaluated the images twice, 6 wk apart, and without knowing which one of the two renograms had been obtained after captopril. The data from this evaluation session were used to assess intra- and interobserver agreement. Additionally, the readers also had to state which one of the two renograms was the one obtained after captopril.

During the second round, they were asked to evaluate the same

Received Aug. 23, 1994; revision accepted Mar. 6, 1995.

For correspondence or reprints contact: P.W. de Leeuw, MD, Department of Internal Medicine, University Hospital Maastricht, P.O. Box 5800, 6202 AZ Maastricht, The Netherlands.

set of renograms paired-wise while the pre- and post-captopril images were indicated as such. The data from this second evaluation session were used to assess the accuracy of the renographic diagnoses in relation to those based on the results of renal angiography. This was done because, in routine clinical practice, nuclear medicine physicians will know which renogram is obtained at baseline and which after captopril. Nevertheless, we used the data from both evaluation sessions to assess whether blinding the readers had any affect on their diagnostic accuracy.

Renogram Evaluation

For the renogram evaluation, the nuclear medicine physicians had at their disposal the sequential renographic images, renographic curves and the value of such variables as fractional uptake, time intervals, total counts and integrated counts at different time intervals. Their conclusions were based on visual interpretation as well as on three well-established criteria for diagnosing RAS as formulated for a post-captopril DTPA renogram (*5*). In the present study, these latter criteria were adopted for ^{99m}Tc-MAG3 renography, that is:

- Percent uptake of ^{99m}Tc-MAG3 by one kidney less than 40% of the total uptake.
- Delayed time-to-peak uptake of ^{99m}Tc-MAG3 in the affected kidney by more than 5 min.
- Delayed excretion of ^{99m}Tc-MAG3 with retention at 15 min as a fraction of peak activity by more than 20% as compared to the contralateral kidney.

In each instance, the readers reported their diagnosis on a three-point scale on which they had to state the probability of RAS being present: low (<50%), intermediate (50%–75%) or high (≥75%). No objective criteria were formulated for these three categories. If, in their opinion, this probability was greater than 50% (intermediate and high probability), they also had to indicate on which side they thought the stenosis to be present.

When a patient had no angiographic abnormalities, the correct renographic diagnosis had to be less than 50% (low) probability of RAS being present. When a unilateral or bilateral stenosis was found on angiography, the correct renographic diagnosis was the one that indicated an intermediate or a high probability for the presence of RAS. Furthermore, in unilateral stenosis, the correct side had to be indicated. For bilateral stenoses on the angiogram, the correct renographic diagnosis was either bilateral or unilateral stenosis.

Statistical Analysis

Calculations were made for observed agreement and kappa (observed agreement corrected for chance agreement) as parameters for both intra- and interobserver variability (*6*). Interobserver variability was calculated on the basis of the presumed probability level for the presence of RAS as indicated by the readers on the three-point scale (low, intermediate and high).

To determine intraobserver variability, the results of the two evaluations of renograms during the first round were compared; intraobserver variability was again calculated in terms of observed agreement and kappa. Sensitivity, specificity, positive and negative predictive values and likelihood ratios were calculated according to accepted methodology (*6*).

Likelihood ratios were calculated only for probability levels ≥50% (intermediate and high), as indicated by the reader. The positive and negative likelihood ratios of each individual physician were added and divided by three and these average values were used to calculate the post-test probability of a positive and negative renogram according to Bayes's theorem (*7*). In this way, a comparison is made between the pre- and post-test probability of angiographically proven RAS in patients with a negative or positive renogram.

TABLE 1
Intraobserver Agreement (OA) and Kappa for Renographic Diagnosis by Three Nuclear Medicine Physicians

Intraobserver agreement	Physician A	Physician B	Physician C
OA for 28 baseline renograms	71%	89%	82%
OA for 28 captopril renograms	64%	79%	79%
OA for all 56 renograms	68%	84%	79%
Kappa for all 56 renograms	0.52	0.75	0.65
OA on type of renogram*	46%	68%	93%

*Intraobserver agreement based on choosing which of the two renograms was the one obtained after captopril. The readers were unaware as to which one of the renograms had been obtained after captopril.

RESULTS

Thirteen patients proved to have RAS (8 unilateral, 5 bilateral), as defined by more than 50% luminal reduction on angiography (*3*), while 15 patients had normal renal arteries and were diagnosed as having essential hypertension.

Intraobserver and Interobserver Variability for Renographic Diagnosis

The data from the first evaluation, in which readers did not know which was the baseline and which the post-captopril renogram, were used to assess intra- and interobserver variability. The results are summarized in Tables 1 and 2.

Intraobserver Agreement (Table 1). Among the three physicians, the intraobserver agreement for baseline and captopril renograms varied from 64% to 89%, with a kappa ranging from 0.52 to 0.75. The data for captopril renograms

TABLE 2
Observed Agreement (OA) and Kappa of Renographic Diagnosis of Technetium-99m-MAG3 Baseline and Captopril Renograms

Baseline renogram	OA Intermed.	OA High	Kappa Intermed.	Kappa High
A vs. B	82%	96%	0.78	0.95
A vs. C	68%	100%	0.61	1.0
B vs. C	86%	96%	0.82	0.95
Captopril renogram				
A vs. B	78%	96%	0.74	0.95
A vs. C	64%	96%	0.57	0.95
B vs. C	75%	93%	0.69	0.91

*Readers had to determine different levels of probability of RAS being present (intermed. = intermediate level, high = high level). The readers were unaware as to which renogram had been obtained after captopril.

did not differ significantly from those for the baseline renograms. When renograms of patients with bilateral stenosis were removed from the analysis, the intraobserver agreement and kappa for baseline and captopril renograms ranged from 35% to 70% and from 0.25 to 0.70, respectively. With respect to the choice of which renogram was the one after captopril, the correctness of choice varied from 71% to 96%, with an intraobserver agreement ranging from 46% to 93%.

Interobserver Agreement (Table 2). Between the three different readers, the observed agreement and kappa ranged from 68% to 86% and from 0.61 to 0.82, respectively, for the evaluation of baseline renograms at an intermediate diagnostic probability level. These figures improved to 96%–100% and 0.95–1.0, respectively, when the reader was convinced of a high probability for the presence of RAS. A similar pattern is seen for the captopril renograms, in which the observed agreement and kappa ranged from 64% to 78% and from 0.57 to 0.74, respectively, at the intermediate probability level and from 93% to 96% and 0.91–0.95, respectively, when the investigator was convinced of a high probability of RAS being present. Results of measurements of interobserver agreement tended to be slightly but not significantly better when renograms of patients with bilateral stenosis were removed from the analysis.

TABLE 3
Different Statistical Parameters of Renographic Diagnosis of Baseline, Captopril and Paired Renography at Intermediate Level of Probability for Presence of RAS in Relation to Renal Angiography

Renography	Physician A	Physician B	Physician C
Baseline*			
Sensitivity	46%	23%	8%
Specificity	80%	93%	93%
PPV	67%	75%	50%
NPV	63%	58%	54%
Captopril*			
Sensitivity	38%	38%	15%
Specificity	67%	87%	93%
PPV	50%	71%	67%
NPV	56%	62%	56%
Paired images†			
Sensitivity	9%	23%	23%
Specificity	47%	40%	33%
PPV	20%	25%	23%
NPV	39%	62%	67%

*Readers were unaware as to which was the baseline or captopril renogram.
†Readers knew which was the baseline or captopril renogram.
PPV = positive predictive value; NPV = negative predictive value.

Renographic versus Angiographic Diagnosis

The results of the evaluation of the paired renograms (baseline plus captopril) in relation to the angiographic diagnoses are presented in Table 3. In this study, the prevalence of RAS proved to be 46%. When renographic diagnoses of the individual physicians are compared with the gold standard (angiography), the results were as follows.

First Evaluation Round (Readers Blinded as to Which of the Two Renograms Was Obtained after Captopril). For renographic diagnoses with an intermediate probability, the sensitivity was less than 50% (range 8%–46%), with a specificity above 65% (range 67%–93%). When only diagnoses with a high probability level were taken into account, sensitivity decreased to values below 20%, with a specificity approaching 100%.

On average, positive and negative predictive values of the paired renograms were 63% and 58%, respectively, at the intermediate probability level; both predictive values improved to values above 75% at a high diagnostic probability level.

Second Evaluation Round (Readers Knew Which of the Two Renograms Was Obtained after Captopril). The average sensitivity (both at an intermediate and a high level of probability) was below 25% (range 5%–25%), with an average specificity below 50% (range 25%–57%). During both evaluation rounds, the three physicians agreed in their renographic diagnoses in 46%–75% of cases. When those cases of complete concordance between the readers were considered separately, the average sensitivity in relation to the renal angiographic diagnosis was below 30% (range 8%–36%), with an average specificity below 60% (range 48%–67%).

In this patient sample, the pre-test probability of angiographically proven RAS is 46%. Given an average calculated positive likelihood ratio of 2.2 during the first (blinded) round and 0.3 during the second round, the post-test probability of a positive renogram is approximately 60% and below 25%, respectively. The post-test probability of a negative test (average likelihood ratio of 1.1 and 2.3) resulted in a post-test probability during both rounds of approximately 45%, which is close to the pre-test probability.

DISCUSSION

In the work-up of hypertensive patients in whom a diagnosis of renovascular hypertension is suspected, captopril renography alone or a paired study (combination with a baseline investigation) is considered to be an essential diagnostic test. No well-established criteria, however, have been formulated for a positive test result (*5,8,9,10*).

Different renographic studies use different criteria to define a positive test or do not define such criteria at all (*11*). Moreover, criteria validated for DTPA renography (*5,8,9*) do not necessarily hold for ^{99m}Tc-MAG3 renography. Apart from renographic patterns (the so called visual interpretation), the most widely used scintigraphic parameters in renography are uptake, time-to-peak, peak activity and residual activity (*5,8,9,10*). Thus, considerable variability may exist in the interpretation of test results and accuracy may not always be guaranteed.

In the present study, we determined the level of agreement between three experienced nuclear medicine physicians with respect to their renographic diagnoses, both in terms of intra- and interobserver variability and as an index of accuracy in relation to the final angiographic diagnoses.

Intraobserver agreement and kappa for the renographic diagnoses ranged from 64% to 89% and from 0.52 to 0.75, respectively, for both baseline and captopril renography. For interobserver variability for baseline and captopril renograms, we found the observed agreement and kappa to range from 68% to 100% and from 0.61–1.0 at different levels of probability. This is in relatively good comparison with other examples in the literature of clinicians interpreting diagnostic tests (*12,13*). In the blinded evaluation, the agreement on the type of renogram (baseline or captopril) varied from 71% to 96%, with an intraobserver agreement ranging from 46% to 93%.

There were no major differences between the three readers according to their diagnostic accuracy and all three performed better when they thought RAS to be present. The overall sensitivity of the physicians' interpretation of the paired renograms, as verified by renal angiography, was rather low and less than 25%. Moreover, sensitivity did not improve even when the three physicians agreed about the diagnosis.

Given a 46% pre-test probability (prevalence) of angiographically proven RAS in our patient population, the post-test probability of a negative renographic study is close to the pre-test probability of 46%. Therefore, the value of a negative renogram in the work-up of these patients is nil. In the event case of a positive renographic study, the results differ depending on whether readers are blinded as to which renogram is the one obtained after captopril. Whereas the post-test probability of RAS being present fell below 25% in the open session, it was increased to 60% in the blinded evaluation. Apparently, nuclear medicine physicians tend to perform better when they do not know which is the baseline and which the captopril renogram. Thus, it may be that a priori knowledge of captopril-induced changes may introduce a diagnostic bias. Although one of the conclusions from this study may be that the diagnostic accuracy of nuclear medicine physicians in renographic studies needs improvement, one could equally argue that the criteria to define a positive or negative ^{99m}Tc-MAG3 renography need to be scrutinized.

Finally, we have to consider that a diagnosis of RAS in a hypertensive patient does not necessarily mean that the patient has renovascular hypertension. In clinical practice, however, captopril renography is being used to screen patients for angiography. Therefore, this study was designed to assess how nuclear medicine physicians differ in renographic diagnoses.

CONCLUSION

The results of our angiographically controlled study of the diagnostic performance of three nuclear medicine physicians in their interpretation of renograms indicate that the intra- and interobserver agreement is reasonably good but that their interpretation of the renogram shows poor sensitivity and post-test probability in comparison to the angiographic diagnosis. Blinding the reader as to which renogram is the pre- and post-captopril image results in better diagnostic accuracy for positive scans. Based on these results, we suggest not performing renographic studies in the work-up of hypertensive patients in whom a diagnosis of renovascular hypertension is already strongly suspected on the basis of clinical clues.

REFERENCES

1. Blackburn H, Blomgvist G, Freiman A, et al. The exercise electrocardiogram: differences in interpretation. Report of a technical group on exercise electrocardiography. *Am J Cardiol* 1968;21:871–877.
2. Chamberlain J, Ginks S, Rogers P. Validity of clinical examination and mammography as screening tests for breast cancer. *Lancet* 1975;2:1026–1028.
3. Bookstein JJ, Abrams HL, Buenger RE, et al. Radiologic aspects of renovascular hypertension. Part 3. Appraisal of arteriography. *JAMA* 1972;221:368–374.
4. Final Report. Working Group on Renovascular Hypertension. *Arch Intern Med* 1987;147:820–829.
5. Mann SJ, Pickering TG, Sos TA, et al. Captopril renography in the diagnosis of renal artery stenosis: accuracy and limitations. *Am J Med* 1991;90:30–40.
6. Armitage P, Berry G. *Statistical methods in medical research*, 2nd ed. Oxford: Blackwell Scientific; 1985.
7. Fagan TJ. Nomogram for Bayes's theorem [Letter]. *N Engl J Med* 1975;293:257.
8. Chen CC, Hoffer PB, Vahjen G, et al. Patients at high risk for renal artery stenosis: a simple method of renal scintigraphic analysis with Tc-99m DTPA and captopril. *Radiology* 1990;176:365–370.
9. Sfakianakis GN, Jaffe DJ, Bourgougnie JJ. Captopril scintigraphy in the diagnosis of renovascular hypertension. *Kidney Int* 1988;34 (suppl):S142–S144.
10. Fommei E, Mezzasalma L, Ghione S, et al. European captopril radionuclide test multicenter study. Preliminary results: inspective renographic analysis. *Am J Hyperten* 1991;4:690S–697S.
11. Prigent A. The diagnosis of renovascular hypertension: the role of captopril renal scintigraphy and related issues. *Eur J Nucl Med* 1993;20:625–644.
12. Zir LM, Miller SW, Dinsmore RE, Gilbert JP, Harthorne JW. Interobserver variability in coronary angiography. *Circulation* 1976;53:627–632.
13. Detre KM, Wright E, Murphy ML, Takaro T. Observer agreement in evaluating coronary angiograms. *Circulation* 1975;52:979–986.