A Software Engine to Justify the Conclusions of an Expert System for Detecting Renal Obstruction on ^{99m}Tc-MAG3 Scans

Ernest V. Garcia, Andrew Taylor, Daya Manatunga, and Russell Folks

Department of Radiology, Emory University School of Medicine, Atlanta, Georgia

The purposes of this study were to describe and evaluate a software engine to justify the conclusions reached by a renal expert system (RENEX) for assessing patients with suspected renal obstruction and to obtain from this evaluation new knowledge that can be incorporated into RENEX to attempt to improve diagnostic performance. **Methods:** RENEX consists of 60 heuristic rules extracted from the rules used by a domain expert to generate the knowledge base and a forward-chaining inference engine to determine obstruction. The justification engine keeps track of the sequence of the rules that are instantiated to reach a conclusion. The interpreter can then request justification by clicking on the specific conclusion. The justification process then reports the English translation of all concatenated rules instantiated to reach that conclusion. The justification engine was evaluated with a prospective group of 60 patients (117 kidneys). After reviewing the standard renal mercaptoacetyltriglycine (MAG3) scans obtained before and after the administration of furosemide, a masked expert determined whether each kidney was obstructed, whether the results were equivocal, or whether the kidney was not obstructed and identified and ranked the main variables associated with each interpretation. Two parameters were then tabulated: the frequency with which the main variables associated with obstruction by the expert were also justified by RENEX and the frequency with which the justification rules provided by RENEX were deemed to be correct by the expert. Only when RENEX and the domain expert agreed on the diagnosis (87 kidneys) were the results used to test the justification. **Results:** RENEX agreed with 91% (184/203) of the rules supplied by the expert for justifying the diagnosis. RENEX provided 103 additional rules justifying the diagnosis; the expert agreed that 102 (99%) were correct, although the rules were considered to be of secondary importance. **Conclusion:** We have described and evaluated a software engine to justify the conclusions of RENEX for detecting renal obstruction with MAG3 renal scans obtained before and after the administration of furosemide. This tool is expected to increase physician confidence in the interpretations provided by RENEX and to assist physicians and trainees in gaining a higher level of expertise.

Key Words: diuresis renography; expert systems; justification engines

J Nucl Med 2007; 48:463–470

The time that a physician has available to devote to individual clinical studies continues to shrink because of increases in procedure volume and related paperwork as well as the time required to assimilate an ever-expanding knowledge base. Moreover, the information that needs to be assimilated to interpret each study continues to increase because of the clinical data and past studies available in expanded digital storage of a patient's records as well as the increased number of images that are acquired because of improved temporal and spatial resolution. Decision support systems have been suggested as an artificial intelligence tool to help physicians interpret diagnostic studies with a high level of expertise while limiting the time needed for interpretation.

We previously developed a renal expert system (RENEX) as a decision support tool to assist in the interpretation of ^{99m}Tc-mercaptoacetyltriglycine (MAG3) scans for detecting renal obstruction in patients referred for diuresis renography (*1*). We chose to develop a decision support system for detecting renal obstruction via ^{99m}Tc-MAG3 renography for 2 reasons: the vast majority of the 590,000 renal scans performed annually in the United States are performed with ^{99m}Tc-MAG3, and many are interpreted by diagnosticians at sites that perform fewer than 3 studies per week (*2*). The exposure to a limited number of diuresis renography studies makes it difficult for physicians to develop the needed diagnostic expertise.

Expert systems for nuclear medicine have been investigated for assisting in the interpretation of perfusion–ventilation lung studies (*3*) and hexamethylpropyleneamine oxime brain SPECT studies (*4*). We have also developed (*5*) and extensively validated (*6*) an expert system called PERFEX (for "perfusion expert") as a tool for computer-assisted diagnosis via stress–rest myocardial perfusion SPECT and PET.

One of the main reasons that we have chosen the expert system approach is that it allows for justification of the

Received May 12, 2006; revision accepted Sep. 5, 2006.
For correspondence or reprints contact: Ernest V. Garcia, PhD, Emory University Hospital, Room E163, 1364 Clifton Rd., NE, Atlanta, GA 30322.
E-mail: Ernest.Garcia@emoryhealthcare.org

conclusions reached by the system. This justification process is important because it provides more experienced diagnosticians with the rationale for the diagnosis, allowing them to agree or disagree with the conclusions on the basis of whether they agree or disagree with the justifications (reasons) for the conclusions. For less experienced (less knowledgeable) diagnosticians, the justification process provides an opportunity for training on a case-by-case basis by teaching them the specific established rules that apply in each case.

A decision support system may provide the right answer for the wrong reasons. The previously described benefits of the justification process are only true if the expert system reaches the correct diagnosis through the correct reasoning process. The purposes of this study were to describe and evaluate a software engine to justify the conclusions reached by RENEX for assessing patients with suspected renal obstruction and to obtain from this evaluation new knowledge that we will eventually incorporate into RENEX to test for improved diagnostic performance.

MATERIALS AND METHODS

The justification engine is a software module of RENEX. Thus, the acquisition protocol and data analysis methods described here are the same as those used for RENEX (*1*). The patient population used for this validation is different from the training group used to develop RENEX and as such represents a prospective population.

Patients

Renal studies from 60 patients (28 men and 32 women; mean age, 53.8 ± 17.6 y; 117 kidneys) were used as a pilot group to test the RENEX justification engine. All studies used for this development were obtained from the renal database of patients referred to our nuclear medicine service for the evaluation of suspected renal obstruction. This study was performed under the purview and approval of the Emory University Internal Review Board. Patients were selected because their studies included a baseline ^{99m}Tc-MAG3 dynamic study followed by a furosemide challenge. Patients were randomly selected from a database of 2,000 patients referred from January 1998 to December 2005 to our nuclear medicine service for the evaluation of suspected renal obstruction.

Acquisition Protocol

Patients were positioned supine, with the scintillation camera detector placed under the table. A 3-phase dynamic acquisition (baseline scan) was begun as approximately 370 MBq (10 mCi) of ^{99m}Tc-MAG3 were injected; phase 1 consisted of twenty-four 2-s frames, phase 2 included sixteen 15-s frames, and phase 3 included forty 30-s frames. For 44 of the patients in the study, review of the baseline scan could not exclude obstruction; these 44 patients received an intravenous injection of approximately 40 mg of furosemide, which was followed immediately by a second, single-phase, 20-min dynamic acquisition consisting of forty 30-s frames. For the other 16 patients, the baseline scan excluded obstruction, and they did not receive furosemide.

Data Analysis

All patient studies were processed with QuantEM version 2.0 software, an improved version of a renal quantification program (*7*). The QuantEM software, developed specifically for ^{99m}Tc-MAG3, incorporates several quality control procedures to improve reproducibility, generates specific quantitative parameters recommended for scan interpretation, and allows MAG3 clearance to be calculated by use of a camera-based technique. The QuantEM software was previously extensively validated in a multicenter trial (*8*).

For the baseline renogram, a static image is summed from the frames obtained at 2–3 min after injection. Through the use of a filtered version of this image, whole-kidney, background, and cortical regions of interest (ROIs) are automatically defined. The user can override any of these automatic ROIs and replace them with manual ROIs. Background-subtracted curves are generated for the whole kidney, and 47 quantitative parameters are generated; these include patient demographics (height, weight, age, sex, and body surface area), curve parameters (time to peak counts and the ratio of counts at 20 min to maximum counts for both whole-kidney and cortical ROIs), voiding indices (ratio of postvoid counts to prevoid counts and ratio of prevoid counts to maximum counts), and MAG3 clearance. MAG3 clearance is calculated from the whole-kidney MAG3 counts at 1–2.5 min and the preinjection and postinjection images of the dose syringe.

For the diuretic study, a static image is summed from the frames obtained at 1–5 min after injection. ROIs are manually drawn for the whole kidney, background, and renal collection system. Background-subtracted curves are generated for the whole kidney and renal pelvis, and times to half-peak counts are calculated.

After the diuretic study is processed, the baseline renogram results are loaded, and the QuantEM software calculates ratios by comparing the first-minute and prevoid (last-minute) counts on the diuretic acquisition with the counts at 1–2 min and peak counts on the baseline acquisition.

RENEX

RENEX was previously described elsewhere (*1*). Briefly, normal limits were established for 47 quantitative parameters extracted from the ^{99m}Tc-MAG3 scans of 100 potential renal donors (*9*). From these data, a domain expert estimated 5 boundary conditions for each parameter: definitely abnormal, probably abnormal, equivocal, probably normal, and definitely normal. A sigmoid-type fit constrained to these 5 boundary conditions was then performed, creating a parameter knowledge library used for converting the value of a prospective patient's individual quantitative parameters to a certainty factor (CF). CF values indicate the degree of certainty that each parameter value is abnormal or normal and, therefore, consistent or inconsistent with disease. CF values range between –1 (definitely normal) and +1 (definitely abnormal), with the interval between –0.2 and +0.2 representing missing, equivocal, or indeterminate parameters. Sixty heuristic rules (if A, then B) were extracted from the rules used by the domain expert to generate the knowledge base for detecting obstruction. A forward-chaining inference engine was developed to determine obstruction. The inference engine is a computer algorithm that uses specific equations known as the MYCIN combinatories (an approximation of Bayes' theorem) to combine the certainty that a parameter (or parameters) is abnormal with the certainty of a rule to modify the certainty that a hypothesis is true (a parameter is abnormal or a kidney is abnormal, that is, obstructed) (*10*).

Expert Review

The domain expert reviewed the standard renal MAG3 scans obtained before and after the administration of furosemide, together with the associated quantitative results determined with the QuantEM version 2.0 software, and reached a diagnosis of obstruction, no obstruction, or equivocal for obstruction. For each kidney diagnosed as obstructed, the masked domain expert identified and ranked the main variables associated with his impression (diagnosis) of obstruction. The domain expert then tabulated 2 parameters: the frequency with which the main variables associated with the diagnosis of obstruction by the expert were also justified (reported as a reason for the diagnosis of obstruction) by RENEX and the frequency with which additional justification rules provided by RENEX were deemed to be correct by the expert. Because RENEX did not have access to the clinical history, this information was not made available to the domain expert. After all studies were independently interpreted by RENEX and the domain expert, a methodical interview session was conducted. This session was similar to the original interview sessions used to generate the RENEX knowledge base, but now the interview was guided by the justification process. The interview took place on a case-by-case basis. For each study, the knowledge engineer requested that the domain expert disclose the interpretation of the study (obstruction or no obstruction) and the ranking of the main variables used to reach the interpretation. After this ranking was recorded, the knowledge engineer disclosed the report generated by the justification engine user interface to the domain expert. Each conclusion reached by RENEX was queried manually by clicking on the specific conclusion in the report by use of the justification engine user interface, which then provided a list of the concatenated rules and variable values used by RENEX. The domain expert then scored whether each of these additional findings was correct or incorrect. The same process was repeated for the studies diagnosed by the domain expert as showing no obstruction.

Only when RENEX and the domain expert agreed on the diagnosis of obstruction or no obstruction were the results used to test the justification. For kidneys for which there was disagreement between RENEX and the domain expert or when both agreed that the study was equivocal for obstruction, the justification was used as a process to discover how to improve the existing rules or add new rules to RENEX.

Justification Engine

The justification engine is a computer algorithm that keeps track of the rules, order of the rules, and CF values for all parameters at the time of firing for all rules that are fired (instantiated). This process is used to justify any conclusion reached by RENEX by simply providing the history of how the conclusion was reached. The justification engine builds and implements a sequential list of the rules that are fired to reach each conclusion and simultaneously tracks the CF value for each pertinent variable as it dynamically changes. Specifically, the engine uses a text array, 1 entry per rule, and a subarray, 1 entry per parameter. When a rule is fired, the English text for that rule is entered into the text array, and the values for the parameters associated with the rule are entered into the subarray. These steps are repeated until a conclusion is reached. The conclusion is as follows: the kidney needs furosemide (Lasix; Aventis Pharmaceuticals, Inc.) or the kidney is obstructed. Once a conclusion is reached, a report is generated (Fig. 1A). The report consists of 2 sections, the findings and the impression. Note that because the justification process uses language common to diagnosticians, the generic name for the diuretic (furosemide) is replaced in the process by the common commercial name Lasix.

Findings. In the findings section, for each kidney, the actual MAG3 clearance and relative uptake parameter values are given, and an interpretation is made as to whether the values are normal, probably normal, equivocal, probably abnormal, or abnormal. This determination is performed by RENEX through conversion of the parameter value to a CF value on the basis of a comparison with the specific boundary conditions for each parameter. Because this interpretation is given by transformations (rather than by rules), only the interpretation is underlined for possible justification. The physician can see the justification for this interpretation, for example, "possibly abnormal," by clicking on the underlined words; subsequently, a parameter window justifying the interpretation appears on the screen and describes the normal value for the specific parameter and the boundary conditions on either side of this value (Fig. 1B).

The findings section also indicates whether or not RENEX concluded that furosemide was needed. This is a rule-based conclusion reached by the expert system rather than a transformation. Clicking on the underlined conclusion that "Lasix was needed" brings up another window that lists the text of all of the concatenated rules that were fired to reach the conclusion that furosemide was needed (Fig. 1C). The parameters associated with each rule are underlined for possible justification. Clicking on a parameter brings up a parameter window as described previously.

Impression. For each kidney, the impression section indicates whether or not RENEX concluded that the kidney was obstructed. Clicking on the underlined conclusion that the "left kidney is probably obstructed" brings up a justification window (Fig. 1D) similar to that described previously for whether furosemide was needed. This window lists the English text of all of the concatenated rules that were fired to reach the conclusion that the kidney was probably obstructed. This justification window now includes the conclusion that furosemide was (or was not) needed as well as all of the other rules; the parameters associated with each rule are underlined for possible justification. Clicking on a parameter brings up a parameter window as described previously.

RESULTS

The results associated with this evaluation include the following: the agreement between RENEX and the domain expert as to whether a kidney was obstructed; the agreement in justifying the same conclusion as to whether a kidney was obstructed, that is, RENEX gave the right answer for the right reasons; and the new knowledge that was discovered as trends when there were equivocal interpretations, disagreement with regard to obstruction, or disagreement with regard to the justification of obstruction and that can be used to improve RENEX.

Justification of Agreements

Figure 2 shows that for 87 kidneys, RENEX and the domain expert agreed; 14 were obstructed, and 73 were not obstructed. For these 87 kidneys, when there was agreement, RENEX agreed with 91% (184/203) of the rules supplied by the domain expert for justifying the diagnosis. RENEX provided 103 additional rules justifying the diagnosis; the expert agreed that 102 (99%) were correct, although the rules were considered of secondary importance. Table 1 provides the frequency and ranking of the

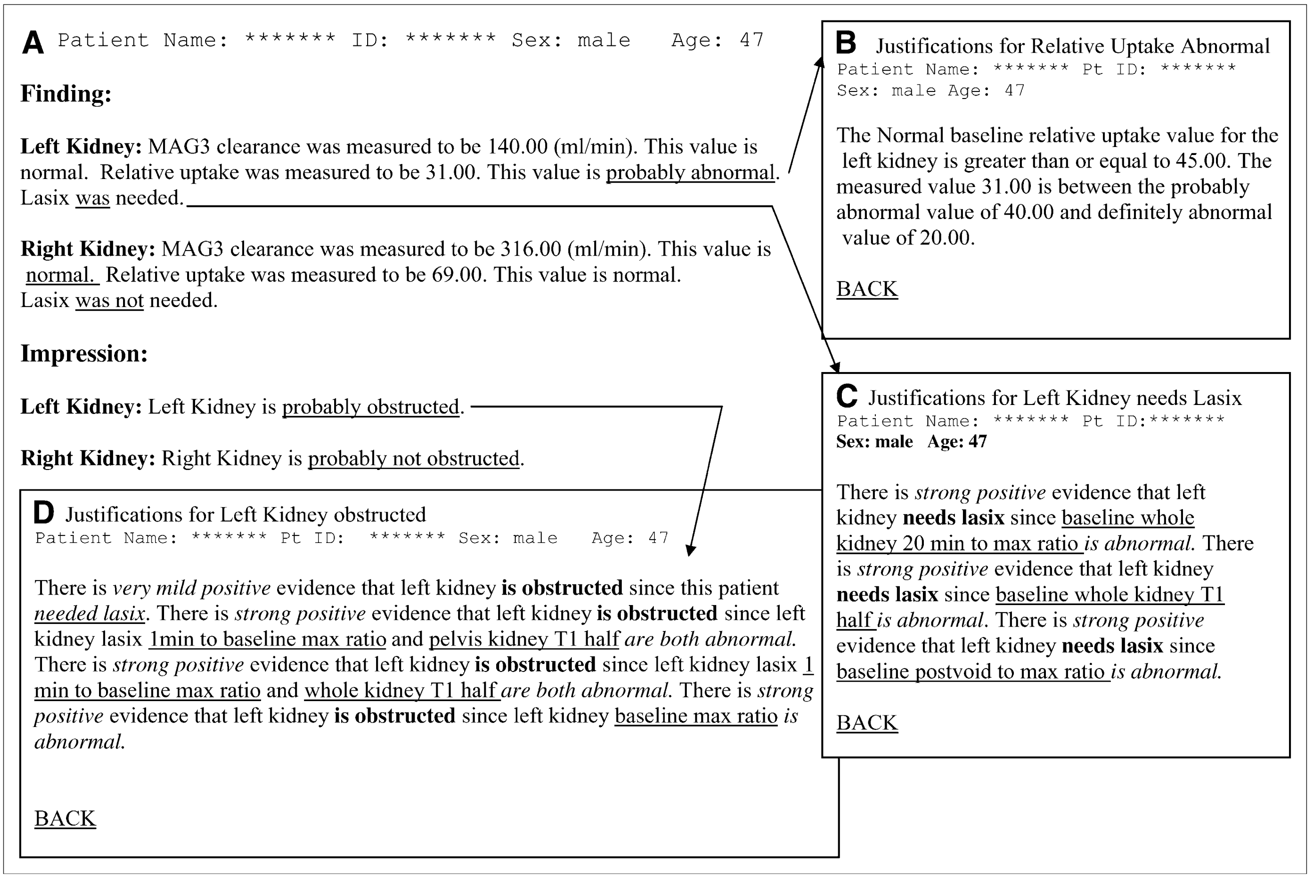

FIGURE 1. Example of RENEX report and justification windows. (A) RENEX report for patient diagnosed as having probably obstructed left kidney and probably not obstructed right kidney. (B) Window providing justification for why relative left kidney uptake was reported as probably abnormal. (C) Window providing explanation for why RENEX concluded that furosemide was needed to determine whether there was obstruction in left kidney. (D) Window providing explanation for why RENEX concluded that left kidney was probably obstructed. max = maximum; min = minimum; T1 half $=T_{1/2}$. See text for more details of this report.

parameters associated with the rules used to justify the presence of obstruction in the 14 kidneys for which there was agreement between RENEX and the expert. Note that the listed parameter with the highest frequency and ranking for this conclusion was the ratio of counts from the kidneys at 19–20 min after furosemide administration (prevoid counts) to the maximum counts on the baseline (prefurosemide) acquisition.

	RENEX								
	A. Left kidney			B. Right kidney			C. All kidneys		
EXPERT	Yes	Eqv	No	Yes	Eqv	No	Yes	Eqv	No
Yes	9	3	0	5	0	0	14	3	0
Eqv	7	4	0	0	3	1	7	7	1
No	1	0	33	4	7	40	5	7	73

FIGURE 2. RENEX vs. expert agreement for determination of renal obstruction. This comparison is shown for left kidneys (A), right kidneys (B), and all kidneys combined (C). Note that for 94 of 117 kidneys (80%), agreement existed between RENEX and expert with regard to obstruction. Eqv = equivocal interpretation; No = no obstruction; Yes = obstruction.

Figure 3 further categorizes the 73 kidneys for which there was agreement that there was no obstruction. For these 73 kidneys, there was agreement that 43 did not need furosemide and were not obstructed and that 19 did need furosemide but were not obstructed, and there were 11 for which RENEX concluded that furosemide was needed but for which the domain expert concluded that furosemide was not needed to exclude obstruction. Table 2 provides the frequency and ranking of the parameters associated with the rules used to justify the conclusion that 43 kidneys did not need furosemide and were not obstructed. Note that the parameter listed with the highest frequency and ranking for this conclusion was the ratio of the postvoid kidney counts to the maximum kidney counts on the baseline acquisition. Table 3 provides the frequency and ranking of the parameters associated with the rules used to justify the conclusion that 19 kidneys did need furosemide but were not obstructed. Note that the parameter listed with the highest frequency and ranking for this conclusion was the ratio of kidney counts from the frame obtained at 19–20 min of the acquisition obtained after furosemide administration (prevoid counts) to the maximum kidney counts on the baseline

TABLE 1
Frequency and Ranking of Rules for RENEX and Expert Agreement in Determining That Furosemide Was Needed and That Kidney Was Obstructed

Kidney	Parameter	Frequency (plotted by rank)	*n*
Left (n = 9)	Furosemide prevoid–to–baseline maximum ratio	111111112	9
	Furosemide renal pelvis $T_{1/2}$	122222233	9
	Furosemide 1 min–to–baseline maximum ratio	2233	4
	Baseline 1–2.5 min relative uptake	334	3
Right (n = 5)	Furosemide prevoid–to–baseline maximum ratio	11111	5
	Furosemide renal pelvis $T_{1/2}$	2223	4
	Furosemide 1 min–to–baseline maximum ratio	23	2
	Baseline MAG3 clearance	34	2
	Furosemide whole-kidney $T_{1/2}$	2	1

(prefurosemide) acquisition. This finding is similar to the finding in Table 1, except that in Table 1 the parameter was found to be abnormal and in Table 3 the parameter was found to be normal.

Trends in Disagreements

A common trend was found for the 11 kidneys in Figure 3 for which both RENEX and the domain expert agreed that the kidneys were not obstructed but for which RENEX concluded that furosemide was needed and the expert disagreed. In all cases, the following 2 rules were fired by RENEX: there was strong positive evidence that the right kidney needed furosemide because the baseline whole-kidney time to half-maximum ($T_{1/2}$) counts was abnormal, and there was moderate negative evidence that the right kidney needed furosemide because the baseline ratio of postvoid counts to maximum counts was normal. Because the negative evidence rule has a weaker qualifier (moderate) than the positive evidence rule (strong), these kidneys were interpreted as needing furosemide. The new knowledge is that these qualifiers should be switched to match the conclusion and diagnostic process of the domain expert.

Figure 2C shows that there was disagreement for 3 kidneys for which the domain expert interpretation was obstruction but for which the RENEX interpretation was equivocal for obstruction. For these kidneys, the justification yielded similar numbers of rules with positive evidence and negative evidence, resulting in the equivocal finding by RENEX. Case-by-case analysis by the domain expert disclosed that because these patients had kidney function (MAG3 clearance) in the normal range, the expert had expected faster MAG3 washout than actually occurred, and this observation increased the expert's certainty that the kidney was obstructed. This observation in various forms generated the new knowledge that the certainty of the MAG3 washout variables should be a function of MAG3 clearance rather than a constant value independent of renal function.

This same trend, but in reverse, was observed for the 7 kidneys in Figure 2C for which the domain expert interpretation was equivocal for obstruction and for the 5 kidneys for which the domain expert interpretation was no obstruction but for which RENEX concluded that all 12 kidneys were obstructed. Here the domain expert concluded that the reduced MAG3 washout could be attributable to impaired (reduced) kidney function measured in terms of absolute MAG3 clearance or relative clearance and, therefore, that the slow washout was not necessarily attributable to obstruction. The expert also believed that these patients with poor kidney function should have received a higher dose of furosemide to adequately challenge these kidneys (*11,12*). Thus, the expert either equivocated with regard to obstruction or interpreted the kidney as not obstructed. Although there is a similar rule in RENEX, the rule tests only for abnormal global MAG3 clearance and not for a reduction in MAG3 clearance for an individual kidney. Similar rules need to be added for borderline abnormal or equivocal MAG3 clearance as well as for borderline abnormal or abnormal relative function to agree with the expert. Moreover, the conclusion as to whether furosemide is needed should be modified to specify the furosemide dose on the basis of renal function or MAG3 clearance.

Figure 2C shows that there was disagreement for 1 kidney for which the domain expert interpretation was equivocal for obstruction but for which the RENEX interpretation was no obstruction. Analysis of this case disclosed that the domain

	RENEX					
	A. Left kidney		B. Right kidney		C. All kidneys	
EXPERT	Yes	No	Yes	No	Yes	No
Yes	3	0	11	0	19	0
No	3	22	8	21	11	43

FIGURE 3. RENEX vs. expert agreement with regard to need for furosemide challenge in kidneys interpreted as not obstructed. This figure represents more detailed analysis of 73 kidneys shown in Figure 2C and interpreted by both RENEX and domain expert as not obstructed. (A) Left kidneys. (B) Right kidneys. (C) All kidneys combined. Note that for 62 of 73 kidneys (85%), agreement existed between RENEX and expert with regard to whether furosemide was needed. No = furosemide not needed; Yes = furosemide needed.

TABLE 2
Frequency and Ranking of Rules for RENEX and Expert Agreement in Determining That Furosemide Was Not Needed and That Kidney Was Not Obstructed

Kidney	Parameter	Frequency (plotted by rank)	*n*
Left (*n* = 22)	Baseline postvoid–to–maximum ratio	1111111111111111111222	22
	Baseline whole-kidney $T_{1/2}$	11122222222222222	17
	Baseline whole-kidney 20 min–to–maximum ratio	23333333	8
Right (*n* = 21)	Baseline postvoid–to–maximum ratio	111111111111111111111	21
	Baseline whole-kidney $T_{1/2}$	222222222222222223	18
	Baseline whole-kidney 20 min–to–maximum ratio	23333	5
	Baseline whole-kidney time to maximum counts	2	1
	Baseline cortical 20 min–to–maximum ratio	3	1
	Baseline MAG3 clearance	4	1

expert noted that this patient had gotten off the table during the acquisition and therefore that the conclusion of no obstruction reached by RENEX because of a large degree of washout was artificial. This realization has pointed out the need for motion detection and correction to be incorporated into subsequent versions of the QuantEM software.

Figure 2C shows that there was disagreement for 7 kidneys for which the domain expert interpretation was no obstruction but for which the RENEX interpretation was equivocal for obstruction. This group did not show a clear trend compared with the other groups, but a case-by-case interview with the domain expert yielded discrepancies similar to those previously described. One additional rule that the expert used but that is not incorporated into RENEX is related to the uniformity of the kidney uptake of MAG3. If the initial uptake in the kidney is uniform without retention in the collection system and washout is present but slow and the images obtained after the administration of furosemide show that uptake continues to be uniform and that MAG3 continues to wash out of the kidney without retention in the collection system, then the slow washout is probably attributable to poor function rather than obstruction.

The last group in Figure 2C to be analyzed was one for which there was actual agreement between the domain expert and RENEX, but the agreement was an equivocal interpretation as to whether the 7 kidneys were obstructed. In general, these were all kidneys with poor MAG3 clearance and for which, according to the domain expert, a higher dose of furosemide should have been used. Beyond identifying the need for more furosemide in these cases, the expert believed that the interpretation of most of these cases would have been facilitated by the availability of the patient's clinical history.

Importantly, for the 3 kidneys for which the domain expert interpretation was obstruction and the RENEX conclusion was equivocal, the CF value for the conclusion was positive (but less than 0.2); in addition, for the 7 kidneys for which the domain expert interpretation was no obstruction and the RENEX conclusion was equivocal, the CF value was negative (but greater than –0.2). These trends show promise for improved accuracy with the knowledge base modifications described here because the conclusions are already leaning in the right direction but fail to reach a diagnostic level of confidence.

DISCUSSION

In this study, we have described how a software engine works to justify conclusions reached by a renal expert system (RENEX) as a decision support tool to assist in the interpretation of ^{99m}Tc-MAG3 scans for the detection of renal obstruction in patients referred for diuresis renography. In kidneys for which there was agreement between RENEX and the domain expert as to whether or not they were obstructed, RENEX agreed with 91% (184/203) of the

TABLE 3
Frequency and Ranking of Rules for RENEX and Expert Agreement in Determining That Furosemide Was Needed but That Kidney Was Not Obstructed

Kidney	Parameter	Frequency (plotted by rank)	*n*
Left (*n* = 8)	Furosemide prevoid–to–baseline maximum ratio	111111	6
	Furosemide 1 min–to–baseline maximum ratio	112222	6
	Furosemide renal pelvis $T_{1/2}$	222	3
Right (*n* = 11)	Furosemide prevoid–to–baseline maximum ratio	1111111122	10
	Furosemide renal pelvis $T_{1/2}$	112222	6
	Furosemide 1 min–to–baseline maximum ratio	22233	5
	Baseline MAG3 clearance	3	1
	Furosemide whole-kidney $T_{1/2}$	3	1

rules supplied by the expert for justifying the diagnosis. We determined that the parameter listed with the highest frequency and ranking for the presence of obstruction was the ratio of the counts from the kidneys in the prevoid frame (19–20 min) of the acquisition obtained after furosemide administration to the maximum counts on the baseline (prefurosemide) acquisition. We also determined that the parameter listed with the highest frequency and ranking for the absence of obstruction was the ratio of the counts from the kidneys in the postvoid frame of the baseline (prefurosemide) acquisition to the maximum counts from the kidneys on the baseline (prefurosemide) acquisition.

Although the $T_{1/2}$ is often considered to be an important quantitative factor in the evaluation of suspected obstruction, the ratios of postvoid counts to maximum counts on the baseline acquisition and the ratios of prevoid counts to maximum counts on the acquisition obtained after furosemide administration were ranked as more important variables by the domain expert. This observation is similar to that reported by others (*13,14*). With regard to the $T_{1/2}$, it is interesting that the domain expert considered that the $T_{1/2}$ based on an ROI encompassing the whole kidney was the most important $T_{1/2}$ measurement for excluding obstruction, whereas the $T_{1/2}$ based on an ROI limited to the counts in the dilated collection system was more important in making the diagnosis of obstruction.

Through the use of the justification engine in a case-by-case analysis of the kidneys for which there was disagreement between RENEX and the domain expert with regard to obstruction, the interview process between the knowledge engineer and the domain expert identified trends in differences in how the rules of the knowledge base were applied, leading to new knowledge that could be incorporated into RENEX to potentially improve performance. Examples of this new knowledge include the following: changes should be made in the degree of certainty of the assertion of the rules concerning the baseline whole-kidney $T_{1/2}$ and the ratios of baseline postvoid counts to maximum counts; the certainty of the MAG3 washout variables should be adjusted so that it is a function of MAG3 clearance (renal function) rather than an expected constant value; patients with poor kidney function should receive a higher dose of furosemide to optimize the diuretic response (*11*); motion detection and correction need to be incorporated into the RENEX algorithms; if the initial uptake in the kidney is uniform and washes out slowly but without pelvic retention and if uptake after furosemide administration continues to be uniform and to wash out from the kidney still without pelvic retention, then the slow washout is probably attributable to poor function rather than obstruction; and the incorporation of clinical history may increase the certainty of the conclusions and yield fewer equivocal findings. We have not yet incorporated the new knowledge and reevaluated the system. This is an iterative process that we plan to continue in the future to attempt to improve the diagnostic performance of RENEX.

A limitation of this evaluation is that only 1 domain expert was used; therefore, the results are dependent on a single person's knowledge. In the past, we have used the consensus interpretation of at least 3 experts to validate an expert system. However, it is important to note that this study is not a validation of RENEX for detecting obstruction but a validation of the RENEX software engine for justifying its conclusions. The analysis was complex enough in evaluating on a case-by-case, rule-by-rule, parameter-by-parameter basis the similarities and differences in how RENEX and the domain expert reached their conclusions. It would have been significantly more difficult to perform this analysis if the comparisons involved various experts. Another limitation is that the text used to justify the conclusions reached does not read smoothly, as though written by a person, but is more a listing of the rules used to reach the conclusions. For the purposes of the analysis in this study, we preferred to have a clear demarcation of rules and were less concerned with the flow of words. We plan to improve this aspect of the user interface as we have done for other expert systems (*6*). In those approaches, we have improved the flow of words by combining similar sentences (or roots of sentences) into 1 sentence, combining sentences leading to 1 conclusion into 1 paragraph, and avoiding repeating the same word in a paragraph by replacing it with a synonym.

These results show that this justification engine provides the right answers for the right reasons; therefore, we speculate that justification of the image interpretation will increase physician confidence in the interpretations provided by RENEX as well as assist less experienced physicians and trainees in gaining a higher level of expertise. Nevertheless, the clinical impact of the justification process has yet to be established. The clinical value of this overall approach will be established only when we demonstrate that RENEX improves the diagnostic performance of physicians in assessing renal obstruction. Importantly, the process of the case-by-case comparison of the rules used by the domain expert and those used by RENEX provided a mechanism for learning how to modify existing rules or add new ones; this information will be incorporated into the knowledge base.

The preliminary validation of the justification process reported here will generate significant enhancements to RENEX. In studies in which the domain expert interpretations, variable rankings, and RENEX conclusions and rules match, no new knowledge is gained. In studies in which the 2 interpretations do not agree or they agree but for different reasons, new knowledge is gained. This new knowledge has been extracted in the form of new rules to be implemented with existing variables, new variables that need to be considered and quantified into the parameter list and eventually incorporated into new rules to enhance knowledge variables, and adjustment of the CF values of existing variables and rules.

Also importantly, to our knowledge, this is the first attempt at validating any justification engine. This methodology may now be generalized for any application.

CONCLUSION

We have described and evaluated a software engine to justify the conclusions of a renal expert system for detecting renal obstruction on ^{99m}Tc-MAG3 renal scans obtained before and after the administration of furosemide. This tool is expected to increase physician confidence in the interpretations provided by RENEX and to assist less experienced physicians and trainees in gaining a higher level of expertise. Importantly, the information obtained from this study will be used by the developers to enhance the knowledge base and then to test for improved diagnostic performance of the decision support system.

ACKNOWLEDGMENTS

This work was funded by National Library of Medicine grant R01-LM007595. Some of the authors receive royalties from the sale of the application software QuantEM. The terms of this arrangement have been reviewed and approved by Emory University in accordance with its conflict-of-interest practice.

REFERENCES

1. Garcia EV, Taylor A, Halkar R, Folks R, Krishnan M, Dubovsky E. RENEX: an expert system for the interpretation of ^{99m}Tc-MAG3 scans to detect renal obstruction. *J Nucl Med.* 2006;47:320–329.
2. IMV Medical Information Division. *2003 Nuclear Medicine Census Market Summary Report.* Des Plaines, IL: IMV Medical Information Division; 2004:IV-7–IV-11.
3. Gabor FV, Datz FL, Christian PE. Image analysis and categorization of ventilation–perfusion scans for the diagnosis of pulmonary embolism using an expert system. *J Nucl Med.* 1994;35:797–802.
4. Imran MB, Kawashima R, Sato K, et al. Detection of CBF deficits in neuropsychiatric disorders by an expert system: a ^{99m}Tc-HMPAO brain SPET study using automated image registration. *Nucl Med Commun.* 1999;20:25–32.
5. Ezquerra N, Mullick R, Cooke D, Krawczynska E, Garcia E. PERFEX: an expert system for interpreting 3d myocardial perfusion. *Expert Syst Appl.* 1993;6: 459–468.
6. Garcia EV, Cooke CD, Folks RD, et al. Diagnostic performance of an expert system for the interpretation of myocardial perfusion SPECT studies. *J Nucl Med.* 2001;42:1185–1191.
7. Taylor A Jr, Corrigan PL, Galt J, et al. Measuring technetium-99m-MAG3 clearance with an improved camera-based method. *J Nucl Med.* 1995;36: 1689–1695.
8. Taylor A Jr, Manatunga A, Morton K, et al. Multicenter trial validation of a camera-based method to measure Tc-99m mercaptoacetyltriglycine, or Tc-99m MAG3, clearance. *Radiology.* 1997;204:47–54.
9. Esteves F, Halkar R, Issa M, Grant S, Waldrop S, Taylor A. Tc-99m MAG3 renal studies: normal values and the correlation between the BSA-corrected camera-based MAG3 and the creatinine clearances [abstract]. *J Nucl Med.* 2002; 43(suppl):351P.
10. Shortliffe EH. *Computer-Based Medical Consultations: MYCIN.* Amsterdam, The Netherlands: Elsevier Scientific Publishing Co.; 1976:264.
11. Hunsche A, Press H, Taylor A. Increasing the dose of furosemide in patients with azotemia and suspected obstruction. *Clin Nucl Med.* 2004;29:149–153.
12. Thrall JH, Ziessman HA. *Nuclear Medicine: The Requisites.* St. Louis, MO: Mosby-Year Book, Inc. 1995:300.
13. Wong DC, Rossleigh MA, Farnsworth RH. Diuretic renography with the addition of quantitative gravity-assisted drainage in infants and children. *J Nucl Med.* 2000;41:1030–1036.
14. Piepsz A, Tondeur M, Ham H. NORA: a simple and reliable parameter for estimating renal output with or without furosemide challenge. *Nucl Med Commun.* 2000;21:317–323.