



Published in final edited form as:

Semin Nucl Med. 2008 January ; 38(1): 67–81. doi:10.1053/j.semnuclmed.2007.09.006.

Decision Support Systems in Diuresis Renography

Andrew Taylor, MD^{*}, Amita Manatunga, PhD[†], and Ernest V. Garcia, PhD^{*}

^{*}Department of Radiology, Emory University School of Medicine, Atlanta, GA

[†]Department of Biostatistics, Emory University School of Medicine, Atlanta, GA

Abstract

The volume of diagnostic imaging studies performed in the United States is rapidly increasing resulting from an increase in the number of patients as well as an increase in the volume of studies per patient. Concurrently, the number and complexity of images in each patient data set are also increasing. Nuclear medicine physicians and radiologists are required to master an ever-expanding knowledge base whereas the hours available to master this knowledge base and apply it to specific tasks are steadily shrinking. The convergence of an expanding knowledge base and escalating time constraints increases the likelihood of physician errors. The problem is particularly acute for low-volume studies such as MAG3 diuresis renography where many imagers may have had limited training or experience. To address this problem, renal decision support systems (DSS) are being developed to assist physicians evaluate suspected obstruction in patients referred for diuresis renography. Categories of DSS include neural networks, case-based reasoning, expert systems and statistical systems; RENEX and CART are examples of renal DSS currently in development. RENEX (renal expert) uses a set of rules obtained from human experts to analyze a knowledge base of expanded quantitative parameters obtained from diuresis MAG3 scintigraphy whereas CART (classification and regression tree analysis) is a statistical method that grows and prunes a decision tree based on an analysis of these quantitative parameters in a training data set. RENEX can be queried to provide the reasons for its conclusions. Initial data show that the interpretations provided by RENEX and CART are comparable to the interpretations of a panel of experts blinded to clinical information. This project should serve as a benchmark for the scientific comparison and collaboration of these 2 fields of medical decision-making. Moreover, we anticipate that these DSS will better define the essential interpretative criteria, foster standardized interpretation, teach trainees to better interpret renal scans, enhance diagnostic accuracy and provide a methodology applicable to other diagnostic problems in radiology and medicine.

The number of diagnostic imaging studies performed annually in the United States is rapidly increasing. In part, this increase is the result of an increase in population but an even more important factor is an increase in the number of studies performed for each patient. The volume of imaging procedures for each Medicare beneficiary, for example, has been increasing about 10% per year for the past 5 years.¹ At the same time the volume of imaging studies is increasing, the number and complexity of the images in each patient data set is also increasing. Imagers are required to master an ever-expanding knowledge base whereas the hours available to master this knowledge base and apply it to specific tasks (selecting the most appropriate protocol, quality control, image interpretation, reporting) are steadily shrinking. The convergence of an expanding knowledge base and escalating time constraints

© 2008 Elsevier Inc. All rights reserved.

Address reprint requests to: Andrew Taylor, MD, Nuclear Medicine, Department of Radiology, Emory University Hospital, 1364 Clifton Rd NE, Atlanta, GA 30322. ataylor@emory.edu.

The terms of the arrangement have been reviewed and approved by Emory University in accordance with its conflict-of-interest practice.

increases the likelihood of physician errors. The problem is particularly acute for low-volume studies such as diuresis renography.

For many full-time nuclear medicine physicians, diuresis renography is a low-volume study. Of the estimated 590,000 renal scans performed annually in the United States, many are interpreted by diagnosticians in sites that perform fewer than 3 studies per week² and even full-time nuclear medicine physicians may disagree up to 20% of the time whether or not a kidney is obstructed, indeterminate, or not obstructed.³ Most nuclear medicine studies in the United States, however, are performed by radiologists. Nuclear medicine is only a small component of most radiology practices, and diuresis renography represents only a small component of nuclear medicine procedures. The problem is compounded by the fact that radiology residents are largely trained to interpret images that require a detailed knowledge of anatomy whereas interpretation of diuresis renography studies depends more on an understanding of the pharmacokinetics of the radiopharmaceutical and renal physiology. Moreover, radiology residents typically receive only 3 months of training to cover all of nuclear medicine compared with 3 years of training for nuclear medicine residents. Nuclear medicine physicians or radiologists with limited experience or insufficient training may try to compensate by overrelying on a single parameter such as the $T_{1/2}$ (time to half maximum counts) after furosemide. If the $T_{1/2}$ is longer than 20 minutes, there is a tendency to interpret the kidney as obstructed. The $T_{1/2}$ is prolonged in obstruction but, depending on how and when the measurement is made, it can be prolonged in a normal individual. Among other factors, the $T_{1/2}$ is affected by hydration, underlying renal function, size and compliance of the renal pelvis, bladder distension, patient position, dose of furosemide and by technical factors such as assignment of regions of interest (pelvis or whole kidney), the algorithm to calculate the $T_{1/2}$ (linear, exponential) and the starting and end points for the $T_{1/2}$ calculation. Naïve and uninformed reliance on a single parameter can lead to inappropriate patient management and unnecessary surgery. In summary, physicians in a conventional imaging practice may lack the time and experience to achieve the desired level of competence in low-volume studies such as diuresis renography, where training may be already limited. For these reasons, it is particularly important to develop and implement decision support tools to help physicians interpret low volume studies such as diuresis renography at a faster rate and at a higher level of expertise.

Consensus Reports as Decision Support Tools

To help standardize practice and guide interpretation of renal scans, an international group of experts in renal nuclear medicine has recently published consensus reports on (1) angiotensin converting enzyme inhibition (ACEI) renography for renovascular hypertension,⁴ (2) diuresis renography,⁵ (3) plasma sample clearance measurements,⁶ (4) quality control of quantitative measurements obtained from the renogram,⁷ (5) technical aspects of renal transplant evaluation,⁸ and (6) pediatric renography.⁹ The consensus recommendations for acquisition of the renogram data, the recommended quantitative parameters, and basic interpretative criteria are now generally accepted by experts but recent British surveys have shown that only 49% of full time nuclear medicine practitioners in Britain were even aware that a guideline on renal clearances existed.¹⁰ The situation is undoubtedly worse in the United States, where most renal scans are interpreted by physicians who practice nuclear medicine part time.² Guidelines and consensus reports have been designed to assist physicians perform and interpret renal studies but for the time constrained physician, they may have made interpretation more complex. To assist in scan interpretation, experts and consensus panels have recommended clearance measurements and the measurement of specific renogram parameters such as time to maximum counts, 20-minute to maximum count ratio, postvoid to maximum count ratio and 20-minute to 2- to 3-minute count ratios for cortical and whole kidney regions of interest (ROIs),⁴⁻⁸ but for many

trainees and practicing physicians, these measurements simply represent a bewildering array of numbers; comfort with the technical requirements of the study and the underlying knowledge of when and how to apply these parameters to assist in scan interpretation may be lacking and, most importantly, physicians may not have time to read and assimilate the relevant papers.

Computer-Based Decision Support (Expert) Systems

To minimize physician errors and improve patient outcomes, tools need to be developed and implemented that will assist physicians in interpreting studies at a faster rate and at a greater level of expertise. Such tools will also minimize subjectivity and intra- and interobserver variation in image interpretation and help achieve a standardized high level of performance. Because diagnostic imaging has become largely digital, computers are a necessary part of acquiring and processing imaging studies and it is reasonable to expect that these new tools should be computer based. During the past several years, artificial intelligence methods have been investigated as a way to develop such tools. Examples include neural networks¹¹⁻¹⁸ and case-based reasoning¹⁹ techniques to provide computer-assisted diagnosis of planar and single-photon emission computed tomography (SPECT) myocardial perfusion studies.¹¹⁻¹⁹ In the artificial neural net approach, the concept is to try to emulate how human neurons perform pattern recognition tasks. For example, repeated recognition trials can be run using sample myocardial perfusion data as input and corresponding coronary angiography results as output to modify the strength between the input and output nodes. In this manner, the net is trained and the input data eventually predicts the output. In the case-based reasoning approach the algorithm searches a library of patient cases to find the ones that best match those of the patient study being analyzed. The common findings from these cases, such as coronary angiography results, are then used to assist the diagnostician's interpretation. Another artificial intelligence approach to assist diagnosticians in making clinical interpretations is the knowledge-based expert system. In expert systems, a knowledge base of heuristic rules is obtained from human experts capturing how they make their interpretations. These rules are usually expressed in the form of "IF A THEN B" expressions.

Expert systems have been investigated in nuclear medicine to assist in the interpretation of perfusion-ventilation lung studies,²⁰ captopril renography,^{21,22} hemamethylpropyleneamine oxime brain SPECT studies,²³ and stress/rest myocardial perfusion SPECT.²⁴⁻²⁶ Using our previous expertise,^{25,26} we have developed a generalized methodology to aid in the interpretation of imaging studies using an expert system to analyze quantitative data extracted from imaging studies, and have applied this generalized methodology to develop a renal expert system (RENEX) for detecting renal obstruction using pre and post furosemide ^{99m}Tc mercaptoacetyltriglycine (MAG3) renal scans.

Why Have We Chosen to Develop Both an Expert System and a Statistical Predictive Model for Diuresis Renography?

Two primary issues of concern for any decision support system are its accuracy and its clinical acceptance. Both are necessary in terms of the ultimate clinical utility. Statistical methods for prediction incorporate scientific knowledge about the data and they incorporate the variability of the measurements in the observed data. They also have the ability to make inferences from a sample to a population including hypothesis testing and estimation of parameters of interest such as misclassification rates. Direct application of standard statistical methods such as logistic regression etc. use well-understood mathematical techniques; however, standard statistical programs have 3 failings that are impediments to acceptance by physicians in that they have no real "understanding" of their problem area,

they have no mechanism for “discussing” their knowledge with the user, they have no means for “explaining” (justify their findings) to physicians.²⁷ More sophisticated statistical methods, however, can overcome some of these deficiencies.

Knowledge-based systems, with their emphasis on knowledge representation, offer a natural environment for implementing the tools that are lacking in many statistical approaches and can provide a rationale for medical decisions.²⁸ Nevertheless, knowledge-based systems may implement statistical techniques that can benefit from the development of more formal mathematical approaches. We have chosen to implement predictive statistical modeling and a heuristic expert system with the expectation that each approach will inform and strengthen the other approach. Moreover, we hope this effort will serve as a benchmark for the scientific comparison and collaboration of these two important fields of medical decision-making.

The Gold Standard for Diuresis Renography: Expert Interpretations or Outcome?

Use of clinical outcome as the gold standard for a diuresis renography decision support system is an attractive goal, but it misses the point of an expert system, which is to interpret studies with the same level of expertise as experts. It is generally accepted that experts interpret studies in their imaging specialty better than general radiologists; this is the basis for having distinct areas of expertise within academic radiology departments and private practice settings. Outcome is certainly an important measure but, in diuresis renography, outcome as a gold standard is confounded by the fact the scan interpretation (obstruction versus no obstruction) has a major impact on the clinical outcome (surgical intervention versus observation); consequently, this gold standard is biased. An additional problem is illustrated by a patient who had a pyeloplasty to relieve obstruction 1 year after a diuresis renography scan was interpreted as “no obstruction.” In this example, did the scan miss obstruction, was the study interpreted incorrectly, did the patient only become obstructed one year following the scan or did an aggressive surgeon operate on a nonobstructed kidney? Using patient outcome as a gold standard has an inherent bias, interpretation of the results is not straightforward and it is not the goal of an expert system.

Choice of Radiopharmaceutical and Furosemide Protocol

Our protocol is based on the 1996 international consensus panel recommendations for diuresis renography.⁵ We use ^{99m}Tc mercaptoacetyltriglycine (MAG3) because the consensus panel considered it to be diagnostically superior to ^{99m}Tc DTPA. The consensus panel recommended a single 35-minute continuous acquisition (single stage) with furosemide administered at 20 minutes; an alternative protocol was to break the continuous acquisition into two stages with a 20-minute baseline acquisition followed by furosemide administration and a second acquisition. Kuyvenhoven and coworkers have pointed out that the inconvenience of furosemide administration can be omitted if the baseline scan can exclude obstruction²⁹; this approach can reduce medical costs by reducing the camera, computer and technologist time required to complete the furosemide study and physician time required to interpret it. For this reason, we have used the two stage acquisition since 1990 and omitted Stage 2 when the baseline acquisition could exclude obstruction.

Acquisition and Processing Protocols

Patients were hydrated with approximately 10 ounces of water on arrival in the department. Imaging was performed with the patient supine and the scintillation camera detector placed under the table. A three-phase dynamic acquisition was begun at the time of injection of

approximately 10 mCi of ^{99m}Tc MAG3. Phase one consisted of twenty-four 2-second frames, phase two was sixteen 15-second frames, and phase three was forty 30-second frames. Our original processing software, QuantEM 1.0, was developed specifically for ^{99m}Tc MAG3³⁰ and included a camera-based method to calculate the MAG3 clearance, which was validated in a multicenter trial.³¹ QuantEM 1.0 also incorporated several quality control procedures to improve reproducibility, generated specific quantitative parameters recommended for scan interpretation, and was licensed by Emory University to GE Healthcare. To support our expert systems, we have upgraded this acquisition and processing software, now called QuantEM 2.0, and have designed it to automatically perform a more extensive check of quality control, acquire additional input parameters, and transmit these parameters to the decision support systems.^{32,33} The acquisition protocol now consists of two phases: Phase 1 acquires one-hundred twenty 2-second frames followed by the Phase 2 acquisition of eighty 15-second frames. To have the broadest applicability, we have written QuantEM 2.0 in IDL (Interactive Data Language, Research Systems, Inc, Boulder, CO) which can run on a PC or any commercial platform. All patient studies were processed using the QuantEM 2.0 renal quantification program.

To process the baseline renogram, a static image is summed from the 2- to 3-minute postinjection frames. Using a filtered version of this image, whole kidney and cortical ROIs as well as perirenal backgrounds that avoid the ureter and collecting system are automatically defined. The user can override any of these automatic ROIs and replace them with manual ROIs. Background-subtracted whole kidney and cortical curves are generated and 47 quantitative parameters are generated including patient demographics (height, weight, age, sex, body surface area), curve parameters (time to peak counts, and 20 minute to count ratio for both whole kidney and cortical ROIs), voiding indices (postvoid to prevoid and postvoid to maximum count ratios), relative uptake and the MAG3 clearance. The MAG3 clearance is calculated from the 1 to 2.5 minute whole kidney uptake of MAG3 corrected for renal depth and attenuation and the preinjection and postinjection images of the dose syringe.^{30,31,34-36}

The furosemide component of the study is a separate acquisition consisting of forty 30-second frames. Furosemide is administered at the start of the furosemide acquisition; the standard dose of furosemide is 40 mg but the nuclear medicine physician monitoring the study sometimes increases the dose of furosemide to 60 or 80 mg if the MAG3 clearance on the baseline study is reduced or if the patient is known to have an elevated creatinine.³⁷ Technologists approve or modify automatically assigned kidney and background ROIs and assign pelvic ROIs and the time interval for the calculation of the T1/2. Quantitative parameters are automatically extracted from the two acquisitions, placed in an XML file and forwarded to RENEX or the statistical decision support systems for analysis.

What Are the Normal Values for the Camera-Based MAG3 Clearance, Renogram, and Voiding Parameters?

Clearance measurements and other specific quantitative parameters have been recommended to assist in scan interpretation and patient management.^{4-8,38-44} To assist in the interpretation of ACEI renography, for example, the Santa Fe consensus report and the Society of Nuclear Medicine procedure guideline on renovascular hypertension recommend measurements of time to maximum counts (T_{max}) and 20-minute/maximum count ratios for whole kidney and cortical regions of interest.^{4,39} The 20-minute/2- to 3-minute count ratio has been proposed as a useful parameter to simultaneously evaluate clearance and excretion and may be especially useful in monitoring transplant patients to distinguish between acute tubular necrosis and rejection.⁴⁰ A measurement of urine drainage based on a quantitative comparison of postvoid kidney counts to the counts obtained during the prevoid period

improves the sensitivity and specificity for detecting an obstructed kidney.^{41–43} Finally, the postvoid urine volume can easily be determined at the time of the scan and may provide important additional information regarding excretory function.⁴⁴

To develop our decision support systems, we had to specify the normal values for all the parameters we measured. To define the normal ranges for the quantitative parameters and to determine if the normal ranges varied based on age and gender, the archived MAG3 acquisitions from 106 subjects evaluated for kidney donation were processed using QuantEM 2.0.⁴⁵ To summarize the results, the percent relative uptake in the right and left kidneys was 49% and 51% ± 4% respectively; there was no difference between males and females. Cortical values for the time to maximum counts, 20-minute/max ratio and 20-minute/2- to 3-minute ratio were lower than the whole kidney values ($P < 0.001$); the mean cortical 20-minute/max count ratio was 0.19 with a SD of 0.07 and 0.04 for the right and left kidneys, respectively. The mean postvoid/max whole kidney count ratio was <0.1 (Table 1) and the mean postvoid residual bladder volume was <30 mL (Table 2). These results confirm and extend previous studies^{46–48} and establish normal limits adjusted for age and gender.⁴⁵

The mean camera-based MAG3 clearance corrected for body surface area was 321 ± 69 mL/min/1.73 m² (Table 3). Clearance measurements can aid in the interpretation of the renogram and facilitate appropriate patient management.^{6,38,49–52} Plasma sample clearance methods are considered to be superior to camera-based clearances⁶ and MAG3 clearances can be calculated with reasonable accuracy from a single plasma sample obtained 40 to 45 minutes after injection^{51,52}; however, informal surveys indicate that nuclear radiology services in the United States rarely offer plasma sample clearances because of the additional technical expertise required to perform a plasma sample measurement and the necessity of complying with CLIA (Clinical Laboratory Improvement Act) regulations required for in vitro plasma sample clearances. Camera based clearances do not require blood or urine collection and generally provide an acceptable estimate of renal function that is equivalent to or superior to the creatinine clearance.^{53–57} The mean and standard deviation for the BSA corrected camera-based MAG3 clearance (321 ± 69 mL/min/1.73 m²; Table 1) was essentially the same as the plasma sample MAG3 clearance measured in two separate populations of potential renal donors at different institutions, 304 ± 70 and 317 ± 74 mL/min/1.73 m².^{58,59} The camera-based clearance technique used in this study has been validated in a multicenter trial³¹ and an earlier version is commercially available on General Xeleris systems; the camera-based MAG3 clearance is more reproducible than the creatinine clearance.⁵⁷ Other camera-based MAG3 clearance techniques have been described,^{60,61} and some vendors provide software to measure the MAG3 clearance using a camera-based technique similar to the one described here but data comparing their results to a plasma based standard have not been published. Camera-based clearance measurements using software from other vendors should be comparable to those described in Table 1 as long as the programs incorporate similar quality processing and control features (background correction, dose infiltration, avoiding potential dead-time losses, a standardized time zero) and the vendors can provide validation studies to ensure the software is performing as specified.

The Architecture of RENEX

RENEX was inspired by two previously developed expert systems, MYCIN (28 Shot) and PERFEX (perfusion expert; licensed Syntermed, Inc, Atlanta, GA).^{25,26} MYCIN is a pioneering rule-based expert system developed in the 1970s to help physicians determine the appropriate antibiotic for patients with infections; the name “MYCIN” was chosen because many of the available antibiotics included “mycin” in the name of the antibiotic. PERFEX is

a commercially available imaging expert system developed to assist physicians in the interpretation of myocardial perfusion SPECT studies.²⁶ An expert in radionuclide scintigraphy (domain expert) used his experience and the normal limits for the kidney parameters extracted from the ^{99m}Tc MAG3 scans of 106 potential renal donors⁴⁶ to estimate 5 boundary conditions for each parameter: (1) definitely abnormal, (2) probably abnormal, (3) equivocal, (4) probably normal, and (5) definitely normal. A sigmoid-like fit constrained to these 5 boundary conditions was then generated to create a parameter knowledge library to be used for converting the value of any individual quantitative parameter to a certainty factor regarding normality or abnormality (Fig. 1A and B). Certainty factors provide an alternative to conditional probability and can easily be combined to adjust hypotheses as additional evidence becomes available. For example, the certainty factor value of +1 is assigned to indicate that the parameter is “definitely abnormal” and the certainty factor value of -1 is assigned to indicate a parameter is “definitely normal”; the certainty factor to indicate the boundary when an equivocal study becomes probably abnormal is assigned a value of +0.2, a truly equivocal study is assigned a certainty factor of 0 and a certainty factor value of -0.2 is assigned to indicate the boundary when an equivocal value becomes probably normal (Fig. 1A and B). Certainty factors between +0.2 and -0.2 are equivocal; unknown values are also assigned a certainty factor of 0.

Sixty heuristic rules (“IF A THEN B”) were extracted from the domain expert to generate the knowledge base for detecting obstruction; 12 of these 60 rules are specifically applied to the baseline study to determine the need for a furosemide administration. Each rule uses the certainty factors describing the degree or abnormality or normality for each parameter that the rule evaluates to generate a certainty factor regarding the need for furosemide to exclude obstruction. For example, one of the rules states, “If the ratio of the postvoid kidney counts of the postfurosemide renogram to the counts in the baseline renogram during the 1 to 2 minute interval is normal, then there is a very strong evidence (certainty factor of +0.8) that the kidney is obstructed.”

These applied rules are chained together by a forward chaining inference engine. An inference engine is software that selects and executes the rules; the design of the RENEX inference engine follows the MYCIN inference engine by approximating Bayes theorem to combine the certainty factors generated by the relevant rules to reach a conclusion (combined certainty factor) regarding the need for furosemide; the combined certainty factor can range from “definitely needs furosemide (+1.0)” to “definitely does not need furosemide (-1.0).” An example of a meta-rule is one that states that when the combined certainty factor regarding the need for furosemide is in the equivocal range (+0.2 to -0.2), the patient should also receive furosemide. If a kidney does not need furosemide, that kidney is not obstructed.

If furosemide is needed, additional certainty factors are generated for parameters relating to the furosemide acquisition as well as certainty factors for parameters relating values from the furosemide acquisition to the baseline acquisition values such as the ratio of the pre- and postvoid kidney counts of the furosemide acquisition to the maximum kidney counts of the baseline acquisition (Fig. 1). The inference engine then selects and executes the rules to reach a conclusion (combined certainty factor) regarding the presence or absence of obstruction from definitely obstructed (+1.0) to definitely not obstructed (-1.0). Kidneys with combined certainty factors in the equivocal range (+0.2 to -0.2) are indeterminate for obstruction. For example, when the inference engine starts execution, the certainty factor that a kidney is obstructed is 0 (unknown). As production rules are asserted (fired), the certainty factor that the kidney is obstructed increases or decreases based on whether the rule is providing positive or negative evidence that the kidney is obstructed. After all the pertinent rules are asserted (ie, all rules with antecedents ≥ 0.2 are fired), the resulting

certainty factor is the conclusion reached by the inference engine. Thus, if the final certainty factor that the kidney is obstructed is greater than 0.2, the conclusion is that the kidney is obstructed; the larger the certainty factor (closer to the maximum value of 1.0), the greater the confidence that the kidney is obstructed. If the certainty factor is less than -0.2, the kidney is not obstructed, and if it lies between -0.2 and +0.2, the kidney is equivocal for obstruction. These initial rules were modified as the system was trained with patient data. Rules were grouped into knowledge islands to perform 5 functions common for each kidney: (a) consider if furosemide needs to be administered, (b) consider if furosemide does not need to be administered, (c) consider if the kidney is obstructed, (d) consider if the kidney is not obstructed, (e) consider if meta-rules for the kidney applies. Meta rules are rules considered after all of other rules are considered.

A software component called a justification engine was implemented to record the sequence of each rule that was fired and the certainty factor value of all input and output parameters at the time of instantiation to track and justify the logic of the conclusions.³² The justification engine allows a user to query RENEX to determine the rules and parameter values that “justify” or explain the software’s conclusion regarding the need for furosemide.⁶² The architecture of RENEX is summarized in Figure 2.

First Things First: Can the Decision Support Systems Analyze the Baseline Acquisition to Exclude Obstruction?

Our suspected obstruction protocol has been to first obtain a baseline scan. If the baseline scan can exclude obstruction, the furosemide acquisition is omitted. Of 704 renal scans obtained for suspected obstruction from Jan 1994 to July 2002, the baseline examination excluded obstruction in 221 (30%) patients. Consequently, as an intermediate step to develop decision support systems to detect obstruction, we first applied our decision support systems to examine only the baseline parameters to determine whether the baseline scan could exclude obstruction. This choice addressed a clinical problem, only required 12 rules from RENEX and provided the data and experience to develop the more complicated systems needed to analyze a two stage study for the presence of obstruction.

The Statistical Approach (CART)

Decision trees present an attractive way of summarizing expert knowledge for convenient use by nonexperts. Decision trees provide a simple flowchart prescription of a short series of yes/no questions which result in a decision relevant to the scientific question of interest. The statistical use of these concepts was developed in 1984 by Breiman and coworkers⁶³ who named the method “classification and regression trees,” more commonly known by its acronym CART. CART was applied to a training set of 80 randomly selected patients (79 right kidneys and 80 left kidneys) referred for suspected obstruction.⁶⁴ The single decision tree for the right kidney as illustrated in Figure 3.

The basic element of the tree is a node, which can either be an internal node or a terminal node. In Figure 3, internal nodes are represented by circles; terminal nodes are represented by rectangles. A tree is grown in a hierarchical manner. At each internal node, a binary (yes/no) question is asked. As a first step, at the top node, the algorithm finds the parameter among all kidney parameters and the cut-off point among all possible cut-off points that does the best job of differentiating between kidneys in the data set that require and those that do not require furosemide to further evaluate obstruction. This process of splitting at each node is continued until a large tree is constructed. A large tree usually overfits the data (ie, is overly sensitive to irregularities in data). An overfitted tree runs the risk of correctly predicting the outcome for all subjects in the training set, yet ending up so specifically

tailored to the training set that it performs poorly on many other datasets. A pruning rule is thus implemented to determine the proper tree size. As a final step, a misclassification rate is calculated in each terminal node of the tree.

In Figure 3, the tree for the right kidney consists of four levels with 6 terminal nodes. In this particular tree, four kidney parameters were found to be useful: postvoid/max ratio, relative uptake, cortical 20 minute/max ratio, and MAG3 clearance. Whether or not a patient requires furosemide depends on the kidney's values for these variables. For example, a right kidney which has a postvoid/max ratio <0.3781 , relative uptake $<86\%$ and cortical 20 minute/max ratio <0.5084 is predicted not to require furosemide. In the training set, 34 right kidneys had this set of characteristics, and none of them were misclassified. The number 0/34 in the left-most terminal node of Figure 3 indicates the misclassification rate for this particular path. The total misclassification rate was only 2.5% (2/79). However, when this tree was applied to the right kidneys ($n = 64$) in the validation sample, the misclassification rate was 15.62% (10/64). Because the data in the validation set were not used in building the original tree, the increase in the number of misclassified kidneys was expected.

CART With Bagging—The CART algorithm is a commonly used method for building statistical models from simple feature data to predict medical decisions. CART is powerful because it can deal with incomplete data and multiple types of features both in terms of input features and predicted features; moreover, CART produces a tree containing rules that can be easily comprehended (Fig. 3).⁶³ A potential problem with using a single tree (as in standard CART) on which to build a prediction model is that small perturbations in the training data can result in drastically different trees. Errors made in an early split are passed down to subsequent splits, thus compounding the error. To stabilize the algorithm, 1001 classification trees were constructed by the common statistical technique of bootstrapping the training data.⁶⁵ In brief, bootstrap sampling is a process that randomly selects a single kidney from the training set, assigns that kidney to the bootstrap dataset, randomly selects another kidney from the training set (this kidney could potentially be the same as the first kidney), assigns that kidney to the bootstrap dataset and continues this process until a bootstrapped sample the same size as the original training set has been constructed. This whole process was then repeated for 1001 iterations to produce 1001 bootstrapped datasets. A tree (algorithm) was developed for each of the 1001 bootstrapped datasets to determine the need for furosemide. These 1001 trees from the training set were applied to each kidney in the validation data, resulting in 1001 predictions for each kidney regarding the need for furosemide to exclude obstruction. The final prediction regarding furosemide was determined by simple majority vote of the 1001 outcomes. This methodology, called bootstrap aggregation or bagging,⁶⁶ reduces dependence on the training set and stabilizes the prediction algorithm by averaging the results. An odd number of bootstrap samples is chosen to avoid any ties in voting.

The modified CART algorithm with bagging reduced the misclassification rate for the right kidney from 15.62% without bagging to 10.94% ($P = 0.03$). The misclassification rates for these 1001 single trees ranged from 4.69% to 35.94%, indicating large variability for single trees but the bagging misclassification rate for the right kidney was smaller than the mean (and median) misclassification rate of the 1001 bootstrapped samples and bagging had the effect of stabilizing the standard CART analysis. In the prospective data set, CART with bagging accurately predicted the need for furosemide about 90% of the time. A significant disadvantage in the bagging technique is the lack of a simple tree at the end of the procedure on which to base future predictions; the final prediction based on 1001 trees is too complicated to be presented visually. Moreover, CART with bagging cannot provide the interpreting physician with a rationale for reaching a specific conclusion.

An important advantage of the CART algorithm is that it identified and specified the parameters used at the various levels of the bootstrapped sampled trees to determine when obstruction could be excluded without the furosemide acquisition; this analysis provided an important insight into the parameters that are most important in discriminating between obstruction and nonobstruction. The time to half peak ($T_{1/2}$) is frequently cited as an important measurement in evaluating possible obstruction,⁶⁷ but this was *not* an important variable in the CART analysis for determining the need for furosemide (distinguishing between nonobstruction and possible obstruction). In fact, 2 of the 3 most frequently selected parameters at the first level employed a comparison of the counts in the kidney after voiding to an earlier time period (maximum counts or counts at 1–2 minutes; Table 4), and support an earlier study suggesting that voiding indices will provide simple and more robust parameters for evaluating obstruction than the $T_{1/2}$.^{41,68}

CART Versus RENEX

As an intermediate step in the development of our decision support systems to detect obstruction, we conducted a prospective study to compare the decisions regarding the need for furosemide made by a heuristic approach (RENEX) and an analytic approach (CART) with the need for furosemide determined in clinical practice and by expert readers.⁶⁹ Both RENEX and CART used the same pilot group of 31 patients (61 kidneys) as a training set. CART with bagging was applied to construct 1001 classification tree to determine the best separation between kidneys that required furosemide to evaluate obstruction and kidneys that were not obstructed and did not require furosemide. Subsequently, both systems were prospectively applied to 102 patients (200 kidneys) of whom 70 received furosemide; decisions regarding the need for furosemide were compared with the clinical decisions and the decisions of three experts who independently scored each kidney on the need for furosemide and resolved differences by majority vote. RENEX performed better than CART when furosemide was required to further evaluate possible obstruction. RENEX agreed with the experts' decisions to give furosemide in 98% (65/66) of patients whereas CART agreed in 89% (59/66), respectively, $P = 0.03$. In contrast, CART performed better than RENEX when furosemide was not required; CART agreed with the experts' decision to withhold furosemide in 78% of kidneys (87/111) whereas RENEX agreed in only 69% of kidneys (77/111), $P = 0.008$. Both systems can be improved and this study is not sufficient to determine if one approach is inherently superior to the other.^{69,70}

This study was limited by the fact that the training set was relatively small; this limitation was probably more of a disadvantage for a statistical system such as CART than for a heuristic system like RENEX. Knowledge-based expert systems have an advantage over neural nets, case based reasoning, or predictive statistical approaches because development of knowledge-based systems do not require the same large numbers of studies as the other approaches. A second advantage of a knowledge-based system, especially from a learning perspective, is that it is possible to query the system to learn the rules that led to a specific conclusion. For example, RENEX disagreed with the experts in one kidney in regard to the need for furosemide because RENEX gave greater weight to the abnormal $T_{1/2}$ than to the postvoid to maximum count ratio. The experts were not queried but appeared to give greater weight to the images and the low postvoid to maximum count ratio. This interpretation is supported by data from CART, indicating that voiding indices will provide a more robust method for determining the presence or absence of obstruction than the $T_{1/2}$ (Tables 2 and 4).⁶⁴ RENEX can be improved by comparing results and its "reasoning" with expert decisions and adding new rules and/or adjusting the weighting factors. In this case, RENEX should be tested giving greater weight to the postvoid to maximum count ratio than the $T_{1/2}$. Use of RENEX or CART as decision support tools in institutions that employ the baseline plus furosemide protocol has the potential to offer a "second opinion" and help avoid

unnecessary imaging and reduce the technologist, computer, camera, and physician time required to perform the procedure.⁶⁹ Importantly, the results obtained from this study helped in the more complex task of developing decision support systems to actually diagnose or exclude obstruction.

Additional Quality Control Is Needed

A review of the discrepancies between the experts and the decision support systems indicated that additional quality control was needed. Experienced nuclear medicine physicians can sometimes read around errors that affect the quantitative parameters but discrepant results occurred because the decision support systems assumed the quantitative values were correct. One error occurred when a patient got off the table before the study was complete; the 20 minute/maximum count ratio was zero indicating to RENEX complete emptying of the kidney and, therefore, no obstruction but it was obvious to the clinicians that the patient got off the table and that the study was incomplete and nondiagnostic. A reduction in renal function can lead to a delay in drainage of MAG3 from the kidney and RENEX incorporates rules relating the individual MAG3 clearance to the rate of washout. The camera-based MAG3 clearance requires a correction for attenuation based on a regression equation derived from the patient's height and weight.^{34,35} The software can accept data entry in pounds or kilograms but if pounds are entered for kilograms or kilograms entered when the software is expecting pounds, the camera-based MAG3 clearance will be erroneous, the resulting certainty factor describing the MAG3 clearance will be erroneous and a rule may be incorrectly applied.

We have developed software to check the entire technologist input data used by the decision support systems.⁷¹ Checks are made for logical inconsistency (negative and nonnumeric values, impossible clock times, final void time earlier than initial void time, dose counted larger than dose injected) and demographic values outside the expected range. Additional checks flag potentially unreliable results (height and weight outside an expected range, very low time-to-peak kidney counts, infiltrated dose and starting the camera after radiopharmaceutical injection) as well as factors that may lead to unreliable quantitative data such as relative uptake less than 5%; in this setting, the cortical parameters for the poorly functioning kidney may have too much noise to be reliable parameters. To validate the QC software, two technologists not involved in software development processed 83 consecutive clinical studies. QC events were defined as technical (study descriptors that were out of range or were entered and then changed, unusually sized or positioned ROIs, missing frames in the dynamic image set) or clinical (calculated functional values judged likely to be unreliable). Potentially serious QC events were defined as the following: camera started late, significant dose infiltration, left/right side ROIs swapped, background oversubtraction giving a negative renogram curve and missing frames. Technical QC events were identified in 30/83 (36%) studies, clinical QC events were identified in 28/83 (34%) of studies and potentially serious QC events were identified in 5/83 (6%) of studies.⁷¹ This evaluation demonstrates that there are QC issues are not uncommon; they can be identified, flagged, and corrected. If QC issues are recognized but cannot be corrected, the certainty factor associated with the suspect parameter can be modified to have less effect on the final decision.

How Well Does RENEX Work in Detecting Obstruction?

The entire system was fine tuned and tested using a pilot group of 32 patients (63 kidneys) deemed by a panel of 3 experts to have 41 unobstructed kidneys, 13 obstructed and 9 equivocal findings.³² The 32 patient studies used as a training set were selected to try to challenge all branches of the decision tree. As each patient was interpreted by RENEX, the rules and certainty factors were adjusted to match the expert's interpretations. RENEX

agreed with the expert panel in 92% (12/13) of the obstructed kidneys, 93% (38/41) of the unobstructed kidneys, and 78% (7/9) of the kidneys interpreted as equivocal for obstructions.³² Displays of the baseline and furosemide acquisitions used for review by the expert panel are illustrated in Figure 4. Processing time per patient was practically instantaneous using a 3.0 GHz PC programmed using IDL. Although this initial agreement was encouraging, it did not validate the method since the results only applied to patients in the training set.

A second study was performed to test RENEX in a prospective population consisting of 60 randomly selected studies (117 kidneys).⁷² Obstruction was excluded by the baseline scan in 17 subjects; 43 subjects received furosemide followed by a second 20-minute acquisition. An expert and RENEX granted each kidney as obstructed, equivocal and nonobstructed; both the expert and RENEX were blinded to the clinical history. RENEX requested furosemide in 3/17 subjects who did not receive furosemide and whose kidneys were considered by the expert reader to be nonobstructed. Since there was no furosemide study, these studies were assigned an incorrect diagnosis of obstruction by RENEX even though a furosemide acquisition would have probably led RENEX to the correct diagnosis. RENEX agreed with the expert reading in 86% (73/85) of nonobstructed kidneys, 53% (8/15) equivocal kidneys, and in 82% (14/17) of the obstructed kidneys (Table 5). Of the 75 kidneys interpreted as nonobstructed by RENEX, 73 (97%) were interpreted as nonobstructed by the expert reader.

A subsequent and more formal evaluation of RENEX has been performed which consisted only of studies containing both baseline and furosemide acquisitions. Because the baseline scan excludes obstruction in approximately one third of our patients, selecting only studies with both baseline and furosemide acquisitions increased the likelihood of including kidneys considered to be obstructed or equivocal by experts and represented a more challenging population because a large population of clearly nonobstructed kidneys was excluded from study. In this study, 3 experts blinded to clinical information reviewed 95 studies and resolved differences by consensus (Taylor A, Garcia EV, Binongo J, et al, unpublished data, 2007). Their results were compared with RENEX. These results have been submitted for publication and showed that RENEX agreed with the experts as well as the experts agreed with each other.

Does RENEX Get the Right Answers for the Right Reasons?

It is possible that RENEX could give the right answers for the wrong reasons or that RENEX could get the right answer by accident. Knowledge-based decision support systems with a justification engine like RENEX can be queried; the justification engine will respond to the query by providing the rules (reasons) used to reach (justify) a diagnostic decision. To determine if RENEX was giving the right answers for the right reasons, we designed a laborious experiment to validate the complicated process by which an expert physician reaches conclusions as compared with RENEX.⁶² The RENEX justification engine was evaluated in a prospective group of 60 patients (117 kidneys). Validation consisted of a blinded expert reviewing the baseline and postfurosemide MAG3 renal images and quantitative data sets provided by QuantEM 2.0 and then identifying and ranking the main variables used to determine if a kidney is obstructed, equivocal or not obstructed. Two parameters were then tabulated: (1) the frequency the main rules associated with the diagnosis of nonobstruction or obstruction by the expert were also provided by RENEX and (2) the frequency that additional justification rules provided by RENEX were deemed to be correct by the expert. Only kidneys where RENEX and the expert agreed on the diagnosis ($n = 87$) as to the presence or absence of obstruction were used for this evaluation; kidneys indeterminate for obstruction were excluded from analysis. In the 87 kidneys where there was agreement on the diagnosis, RENEX agreed with 91% (184/203) of the rules supplied

by the expert to justify the diagnosis. RENEX provided 103 additional rules justifying the diagnosis and the expert agreed that 102 (99%) were correct although these rules were considered to be of secondary importance. These results show that the justification engine is essentially using the same rules as the expert to reach its conclusions. Importantly, in the cases where there was disagreement, the process of the patient-by-patient comparison between the rules used by the expert and those used by RENEX provide a mechanism for knowledge discovery as to how to modify existing rules or add new ones to improve the performance of RENEX. To our knowledge, this is the first attempt at validating any justification engine. In an invited perspective, Porenta points out that the clinical acceptability of an expert system strongly depends on user acceptance and user acceptance can only be achieved if the user has confidence in and accepts the reasoning process of the expert system.⁷³ Our study documents that our rule based expert system gives the right answer for the right reason and has the potential to be used not only to assist physicians in the

Limitations and Future Directions

This study addressed the diuresis renography protocol recommended by the international consensus report where baseline data are obtained followed by the administration of furosemide and an additional period of imaging.⁵ There are other protocols in which furosemide is given 15 minutes before the radiopharmaceutical, at the same time as the radiopharmaceutical or 5 to 10 minutes later.^{5,43,74,75} Obviously, the systems we describe at present do not apply to these protocols. QuantEM 2.0 cannot detect and correct for patient motion and, at this time, the software cannot distinguish between diffuse retention with slow washout due to impaired function and focal pelvic retention with slow wash-out due to possible obstruction. Robust algorithms to assign the kidney regions of interest, algorithms to detect and correct for motion and algorithms to distinguish between diffuse retention in a kidney and retention in a dilated renal collecting system need to be designed, implemented and tested. One of the most important limitations is the absence of clinical information. In all our studies, both the experts and the decision support systems were blinded to clinical information other than the fact that the reason for the scan was suspected obstruction. Our preliminary data suggest that the addition of clinical data will reduce the number of equivocal or intermediate interpretations by 60 to 70%. Our future plans include incorporating clinical information, adapting the decision support systems to other diuresis renography protocols and to apply this approach to patients with suspected renovascular hypertension.

Acknowledgments

This work was supported by a grant from the National Library of Medicine, R01-LM007595. Two authors (A.T., E.V.G.) receive royalties from the sale of the application software QuantEM 1.0 related to the research described herein.

References

1. Iglehart J. The new era of medical imaging—progress and pitfalls. *N Eng J Med.* 2006; 354:2822–2828.
2. IMV Medical information division. 2003 Nuclear Medicine Census Market Summary Report. Vol. IV. IMV Limited; Des Plaines, IL: 2003. p. 7-11.
3. Hunsche, A. PhD Thesis. Federal University of Rio Grande do Sul; Porto Alegre, Rio Grande do Sul, Brazil: 2006. Value of quantitative data in the interpretation of diuresis renography for suspected urinary tract obstruction.
4. 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. [PubMed: 8917196]

5. O'Reilly P, Aurell M, Britton K, et al. Consensus on diuresis renography for investigating the dilated upper urinary tract. *J Nucl Med.* 1996; 37:1872–1876. [PubMed: 8917195]
6. Blaufox MD, Aurell M, Bubeck B, et al. Report of the radionuclides in nephrourology committee on renal clearance. *J Nucl Med.* 1996; 37:1883–1890. [PubMed: 8917197]
7. Prigent A, Cosgriff P, Gates GF, et al. Consensus report on quality control of quantitative measurements of renal function obtained from the renogram: International consensus committee from the scientific committee of Radionuclides in Nephrourology. *Semin Nucl Med.* 1999; 29:146–159.
8. Dubovsky EV, Russell CD, Bischof-Delaloye A, et al. Report of the radionuclides in nephrourology committee for evaluation of transplanted kidney (review of techniques). *Semin Nucl Med.* 1999; 29:175–188. [PubMed: 10321828]
9. Gordon I, Colarinha P, Fettich J, et al. Guidelines for standard and diuretic renography in children. *Eur J Nucl Med.* 2001; 28:BP21–30. [PubMed: 11315613]
10. Cosgriff, PS.; Stevens, D. Impact of radionuclides in nephrourology guidelines. *Alasbimn J.* 2001. Available at www.alasbimnjournal.cl
11. Fujita H, Katafuchi T, Uehara T, et al. Application of neural network to computer-aided diagnosis of coronary artery disease in myocardial SPECT Bull's-eye images. *J Nucl Med.* 1992; 33:272–276. [PubMed: 1732455]
12. Porenta G, Dorffner G, Kundrat S, et al. Automated interpretation of planar thallium-201-dipyridamole stress-redistribution scintigrams using artificial neural networks. *J Nucl Med.* 1994; 35:2041–2047. [PubMed: 7989989]
13. Hamilton D, Riley PJ, Miola UJ, et al. A feed forward neural network for classification of bull's-eye myocardial perfusion images. *Eur J Nucl Med.* 1995; 22:108–115. [PubMed: 7758496]
14. Lindahl D, Palmer J, Ohlsson M, et al. Automated interpretation of myocardial SPECT perfusion images using artificial neural networks. *J Nucl Med.* 1997; 38:1870–1875. [PubMed: 9430460]
15. Lindahl D, Palmer J, Pettersson J, et al. Scintigraphic diagnosis of coronary artery disease: Myocardial Bull's-eye images contain the important information. *Clin Physiol.* 1998; 18:554–561. [PubMed: 9818161]
16. Lindahl D, Lanke J, Lundin A, et al. Improved classifications of myocardial bull's-eye scintigrams with computer-based decision support system. *J Nucl Med.* 1999; 40:96–101. [PubMed: 9935064]
17. Lindahl D, Palmer J, Edenbrandt L. Myocardial SPET: artificial neural networks describe extent and severity of perfusion defects. *Clin Physiol.* 1999; 19:497–503. [PubMed: 10583343]
18. Lindahl D, Toft J, Hesse B, et al. Scandinavian test of artificial neural network for classification of myocardial perfusion images. *Clin Physiol.* 2000; 20:253–261. [PubMed: 10886256]
19. Haddad M, Adlassnig KP, Porenta G. Feasibility analysis of a case-based reasoning system for automated detection of coronary heart disease from myocardial scintigrams. *Artif Intell Med.* 1997; 9:61–78. [PubMed: 9021059]
20. 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(5):797–802. [PubMed: 8176461]
21. Hamilton D, Ueber JM, Mousa D. Interpretation of captopril transplant renography using a feed forward neural network. *J Nucl Med.* 1996; 37:1649–1652. [PubMed: 8862301]
22. Nielsen M, Granerus G, Ohlsson M, et al. Interpretation of captopril renography using artificial neural networks. *Clin Physiol Funct Imaging.* 2005; 25:293–296. [PubMed: 16117733]
23. 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. *Nuc Med Comm.* 1999; 20:25–32.
24. Porenta G, Dorffner G, Kundrat S, et al. Automated interpretation of planar thallium-201-dipyridamole stress-redistribution scintigrams using artificial neural networks. *J Nucl Med.* 1994; 35:2041–2047. [PubMed: 7989989]
25. Ezquerria N, Mullick R, Cooke D, et al. PERFEX: An expert system for interpreting 3d myocardial perfusion. *Expert Systems with Applications.* 1993; 6:459–468.

26. 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. [PubMed: 11483678]
27. Gorry G. Computer-assisted clinical decision making. *Met Info Med.* 1973; 12:45–51.
28. Shortliffe, EH. *Computer-Based Medical Consultations: MYCIN.* Amsterdam, Netherlands: Elsevier; 1976. p. 264
29. Kuyvenhoven J, Piepsz A, Ham H. When could the administration of furosemide be avoided. *Clin Nucl Med.* 2003; 28:732–737. [PubMed: 12972994]
30. Taylor A, Corrigan P, Galt J, et al. Measuring technetium-99m-MAG3 clearance with an improved camera based method. *J Nucl Med.* 1995; 36:1689–1695. [PubMed: 7658232]
31. Taylor A, Manatunga A, Morton K, et al. Multicenter trial validation of a camera-based method to measure Tc-99m mercaptoacetyl triglycine, or Tc-99m MAG3, clearance. *Radiology.* 1997; 204:47–54. [PubMed: 9205222]
32. Garcia EV, Taylor A, Halkar R, et al. RENEX: An expert system for the interpretation of 99mTc-MAG3 scans to detect renal obstruction. *J Nucl Med.* 2006; 47:320–329. [PubMed: 16455639]
33. Folks RD, Taylor AT, Garcia EV. Development and prospective evaluation of an automated software system for quality control of quantitative Tc-99m MAG3 renal studies. *J Nucl Med Tech.* 2007; 35:27–33.
34. Taylor A, Lewis C, Giacometti A, et al. Improved formulas for the estimation of renal depth in adults. *J Nucl Med.* 1993; 34:1766–1769. [PubMed: 8410296]
35. Taylor A. Formulas to estimate renal depth in adults (letter to the editor). *J Nucl Med.* 1994; 35:2054–2055.
36. Taylor A, Thakore K, Folks R, et al. Background subtraction in technetium-99m-MAG3 renography. *J Nucl Med.* 1997; 38:74–79. [PubMed: 8998155]
37. 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. [PubMed: 15162982]
38. Müller-Suur R, Tidgren B, Lundberg HJ. Effect of captopril on MAG3 clearance in patients with and without renal artery stenosis and after PTR. *Eur J Nucl Med.* 1998; 25:845.
39. Taylor, A.; Blaufox, MD.; Dubovsky, EV., et al. Procedure guideline for the diagnosis of renovascular hypertension. Policy and practice, procedure guidelines. Jun. 2003 Available at: Snm.org
40. Li Y, Russell CD, Palmer-Lawrence J, et al. Quantitation of renal parenchymal retention of technetium-99m-MAG3 in renal transplants. *J Nucl Med.* 1994; 35:846–850. [PubMed: 8176469]
41. Piepsz A, Tondeur M, Ham H. NORA: A simple and reliable parameter for estimating renal output with or without frusemide challenge. *Nucl Med Comm.* 2000; 21:317–323.
42. Gordon I, Mialdea-Fernandez RM, Peters AM. Pelviureteric junction obstruction. The value of a post-micturition view in ^{99m}Tc DTPA diuretic renography. *Br J Urol.* 2000; 61:409–412. [PubMed: 3293688]
43. 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. [PubMed: 10855630]
44. Strauss BS, Blaufox MD. Estimation of residual urine volume and urine flow rates without ureteral catheterization. *J Nucl Med.* 1970; 11:81–84. [PubMed: 5410926]
45. Esteves FP, Taylor A, Manatunga A, et al. Tc-99m MAG3 renography: normal values for camera-based 99m-Tc MAG3 clearance, MAG3 curve parameters, excretory parameters and residual urine volume. *AJR Am J Roentgenol.* 2006; 187:W610–619. [PubMed: 17114514]
46. Klingensmith WC, Briggs DE, Smith WI. Technetium-99m-MAG3 renal studies: Normal range and reproducibility of physiologic parameters as a function of age and sex. *J Nucl Med.* 1994; 35:1612–1617. [PubMed: 7931658]
47. Lin WY, Changlai SP, Kao CH. Normal ranges of renal physiological parameters for technetium-99m mercaptoacetyl triglycine and the influence of age and sex using a camera-based method. *Urol Int.* 1998; 60:11–16. [PubMed: 9519415]

48. Clausen TD, Kanstrup I, Jensen J. Reference values for ^{99m}Tc -MAG3 renography determined in healthy, potential renal donors. *Clin Physiol Func Imaging*. 2002; 22:356–360.
49. Schlegel JU, Jamway SA. Individual renal plasma flow determination in 2 minutes. *J Urol*. 1976; 116:282–285. [PubMed: 785027]
50. Gates GF. Glomerular filtration rate: Estimation from fractional renal acculumulation of Tc-99m DTPA. *AJR Am J Roentgenol*. 1982; 138:565–570. [PubMed: 7039273]
51. Russell CD, Dubovsky EV. Single-sample Tc-99m MAG3 renal clearance in the long-term management of patients with spinal cord injury. *Nucl Med Commun*. 1998; 19:494–498.
52. Bubeck B. Renal clearance determination with one blood sample: Improved accuracy and universal applicability by a new calculation principle. *Semin Nucl Med*. 1993; 23:73–76. [PubMed: 8469997]
53. Chachati A, Meyers A, Godon JP, et al. Rapid method for the measurement of differential renal function: Validation. *J Nucl Med*. 1987; 28:829–836. [PubMed: 3553463]
54. Bocher M, Shrem Y, Tappiser A, et al. Tc-99m mercaptoacetyltriglycine— clearance comparison of camera-assisted methods. *Clin Nucl Med*. 2001; 26:745–750. [PubMed: 11507290]
55. Russell CD, Dubovsky EV. Gates method for FR measurement (letter). *J Nucl Med*. 1986; 27:1373–1374. [PubMed: 3734910]
56. Esteves FP, Halkar RK, Issa MM, et al. Evaluation of renal function: A comparison between the camera-based Tc-99m MAG3 and 24-hour creatinine clearances. *AJR Am J Roentgenol*. 2006; 187:W316–319. [PubMed: 16928911]
57. Halkar R, Taylor A, Manatunga A, et al. Monitoring renal function: A prospective study comparing camera-based Tc-99m mercaptoacetyltriglycine clearance and creatinine clearance. *Urology*. 2007; 69:426–430. [PubMed: 17382136]
58. Russell CD, Taylor AT, Dubovsky EV. Measurement of renal function with technetium-99m-MAG3 in children and adults. *J Nucl Med*. 1996; 37:588–593. [PubMed: 8691246]
59. El-Galley R, Clarke HS, O'Brien DP, et al. Normal parameters for Tc-99m MAG3 renography. *J Nucl Med*. 1998; 39:87P.
60. Inoue Y, Ohtake T, Yokoyama I, et al. Evaluation of renal function from ^{99m}Tc -MAG3 renography without blood sampling. *J Nucl Med*. 1999; 40:793–798. [PubMed: 10319752]
61. Itoh K, Nonomura K, Yamashita T, et al. Quantification of renal function with a count-based gamma method using technetium-99m-MAG3 in children. *J Nucl Med*. 1996; 37:71–75. [PubMed: 8544005]
62. Garcia EV, Taylor A, Manatunga D, et al. A software engine to justify the conclusions of an expert system's conclusions for detecting renal obstruction on Tc-99m MAG3 Scans. *J Nucl Med*. 2007; 48:463–470. [PubMed: 17332625]
63. Breiman, L.; Friedman, J.; Olshen, R., et al. *Classification and Regression Trees*. Wadsworth; Belmont, CA: 1984.
64. Binongo JNG, Taylor A, Manatunga A, et al. Use of Classification and Regression Trees in Diuresis Renography. *Acad Radiol*. 2007; 14:306–311. [PubMed: 17307663]
65. Efron, B.; Tibshirani, R. *An Introduction to the Bootstrap*. Chapman and Hall; London: 1993.
66. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, section 9.2. Springer; New York: 2001. p. 246-250.
67. Thrall, JH.; Ziessman, HA. *Nuclear Medicine: The Requisites*. Mosby-Year Book, Inc; St. Louis, MO: 1995. p. 302
68. Piepsz A, Kuyvenhoven JD, Tondeur M, Ham H. Normalized residual activity: Usual values and robustness of the method. *J Nucl Med*. 2002; 43:33–38. [PubMed: 11801700]
69. Taylor A, Hill A, Binongo J, et al. Evaluation of two diuresis renography decision support systems designed to determine the need for furosemide in patients with suspected obstruction. *AJR Am J Roentgenol*. 2007; 188:1395–1402. [PubMed: 17449788]
70. Binongo JNG, Yuan M, Bao J, et al. Prediction based two-stage modeling. *Joint Stat Meetings Proc*. 2006; 2006:3967–3972.

71. Folks RD, Taylor AT, Garcia EV. Development and prospective evaluation of an automated software system for quality control of quantitative Tc-99m MAG3 renal studies. *J Nucl Med Tech.* 2007; 35:27–33.
72. Taylor A, Garcia EV, Folks R, et al. Preliminary diagnostic performance of RENEX, an expert system to detect renal obstruction using Tc-99m MAG3. *J Nucl Med.* 2006; 47:147P.
73. Porenta G. Being right for the right reason: Better than just being right? *J Nucl Med.* 2007; 48:335–336. [PubMed: 17332608]
74. Sundaram PS, Padma S, Bhat S, Sanjeevan KV, Rahul C. F + 10 diuretic renography protocol is better than F-15 in followup of post pyeloplasty patients. *J Nucl Med.* 2003; 44:359P. [PubMed: 12621001]
75. Sfakianakis GN, Cohen DJ, Braunstein RH, et al. MAG3-F0 scintigraphy in decision making for emergency intervention in renal colic after helical CT positive for a urolith. *J Nucl Med.* 2000; 41:1813–1822. [PubMed: 11079488]

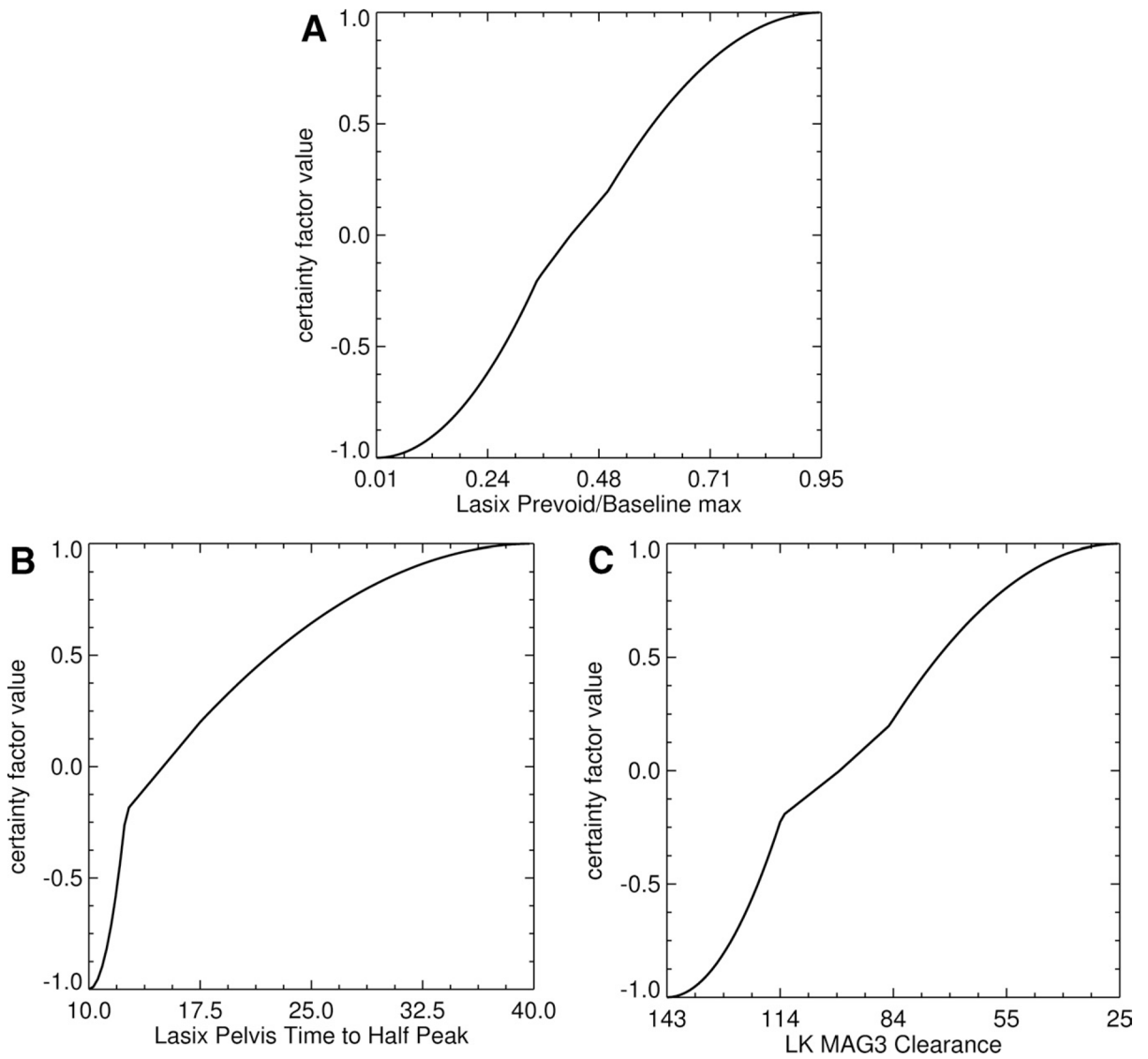


Figure 1. Graphical representation of the transformation of 3 input quantitative parameters to certainty factor values. The 3 parameters are illustrated as follows. (A) Lasix prevoid to baseline max: the ratio of the counts in the kidney ROI during the last frame of the postfurosemide renogram to the maximum counts in the kidney ROI from the prefurosemide baseline renogram. (B) Lasix pelvis time to half peak: the time that it takes for the renogram curve extracted from a kidney's pelvic ROI to decrease from its maximum value to half that value. (C) The MAG3 clearance curve shows the camera-based MAG3 clearance for the left kidney. Notice that the curves have a general sigmoid shape but do not have a smooth, exact, Sigmoid fit. (Reprinted by permission of the Society of Nuclear Medicine from Garcia et al.³²)

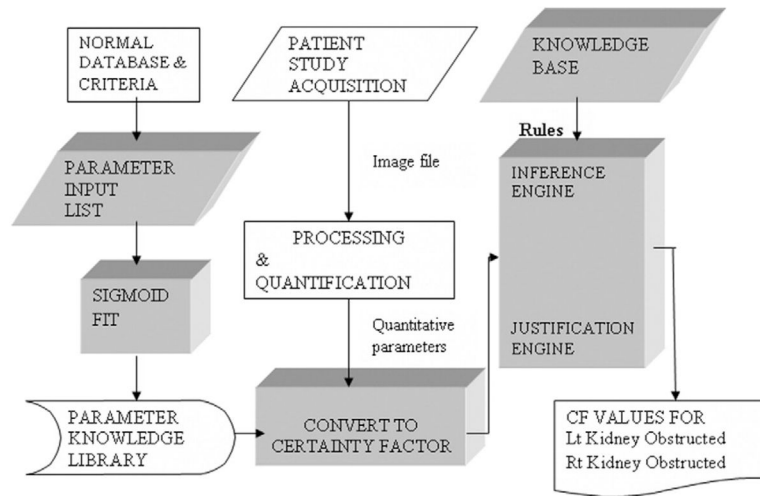


Figure 2.

Flow diagram for RENEX. This diagram shows the flow of how a renal scan is acquired, processed, quantified to extract parameters of renal function and how these parameters are converted to certainty factors (CF) that are then input to the expert system. The expert system is comprised of the knowledge base, the inference engine and the justification engine. The inference engine applies rules from the knowledge base to the certainty factors describing the parameters of the study and combines the certainty factors to reach a conclusion regarding the presence or absence of obstruction. The justification engine keeps track of the order and sequence of the rules that were applied. The trapezoidal blocks indicate domain expert; the rectangular blocks indicate software algorithms. (Reprinted by permission of the Society of Nuclear Medicine from Garcia et al.³²)

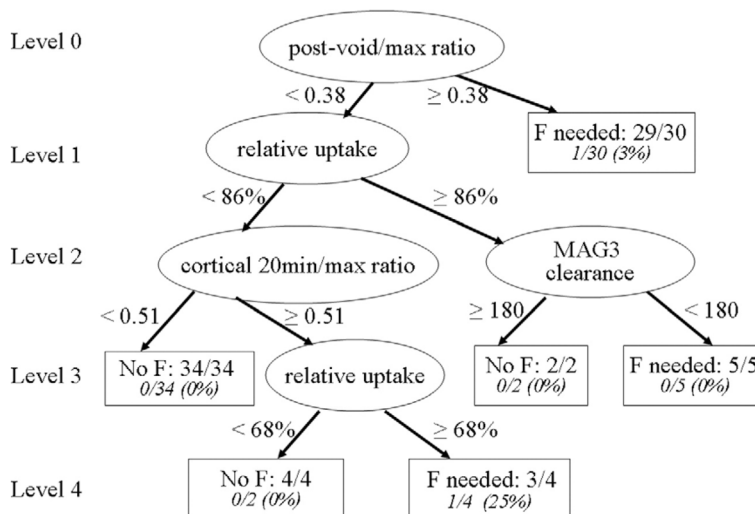
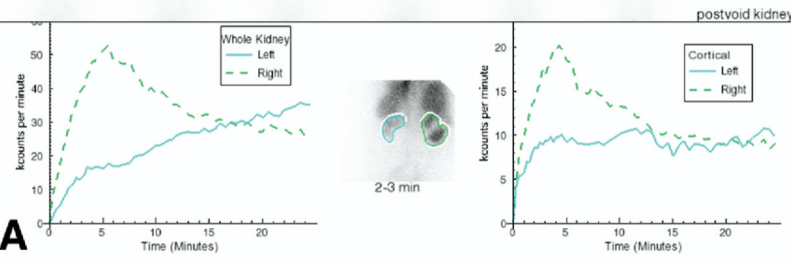
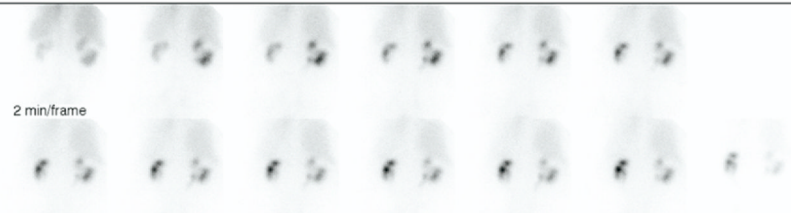


Figure 3. A single decision tree to determine whether furosemide is needed for the right kidney to further evaluate obstruction is illustrated. The basic element of the tree is a node, which can either be an internal node or a terminal node. Internal nodes are represented by circles; terminal nodes are represented by rectangles. At each internal node, a binary (yes/no) question is asked. As a first step, at the top node, the algorithm finds the parameter among all kidney parameters and the cut-off point among all possible cut-off points that does the best job of differentiating between kidneys in the data set that require and those that do not require furosemide to further evaluate obstruction. This tree consists of four levels with 6 terminal nodes; four kidney parameters were found to be useful in determining the need for furosemide: postvoid/max ratio, relative uptake, cortical 20-minute/max ratio, and MAG3 clearance. Whether or not a patient requires furosemide depends on the kidney’s values for these variables. For example, a right kidney which has a postvoid/max ratio <0.3781, relative uptake <86% and cortical 20-minute/max ratio <0.5084 is predicted not to require furosemide. Each terminal node gives the number and percentage of kidneys in that node that were misclassified. (Adapted with permission from Binongo et al.⁶⁴)

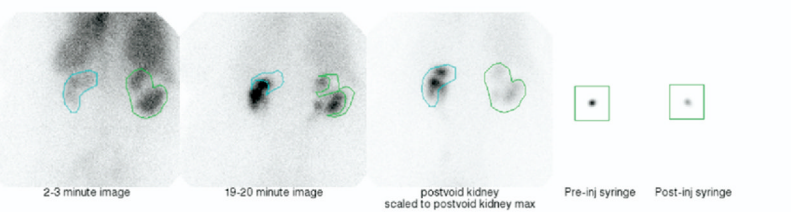
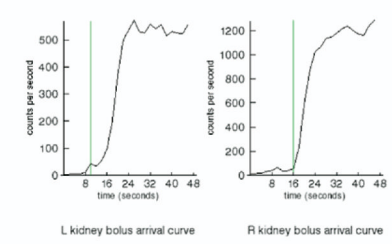
Renogram Review

patient height = 160.0 cm	kidney:	Left	Right
patient weight = 86.0 kg		-----	-----
patient age = 74 y.o. female	% uptake	= 26	74
dose injected = 10.50 mCi	Tmax (min)	= 23.3	5.3
dose counted = 1.10 mCi	TI/2 (min)	= >50	19.0
% infiltrated = 0.05	20/max	= 0.85	0.55
MAG3 clearance = 94 ml/min	voided vol.	= na	
expected clr = (226 - 439)	residual vol.	= na	
	urine flow	= na	



Expanded Renogram Review

patient height = 160.0 cm			
patient weight = 86.0 kg			
patient age = 74 y.o. female			
dose injected = 10.50 mCi			
dose counted = 1.10 mCi			
MAG3 clearance = 94 ml/min	Expected Ranges for female		
	(226 - 439)		
kidney	Left	Right	Left
	-----	-----	-----
% uptake	= 26	74	(45 - 58)
Tmax (min)	= 23.3	5.3	(2.3 - 4.8)
TI/2 (min)	= >50	19.0	(3.7 - 18.5)
20/max	= 0.85	0.55	(0.13 - 0.35)
postvoid/max	= 1.12	0.36	(0.05 - 0.15)
cortical			
	-----	-----	-----
Tmax (min)	= 23.3	4.3	(2.1 - 3.3)
TI/2 (min)	= >50	19.0	(3.8 - 7.3)
20/max	= 0.79	0.47	(0.13 - 0.27)
voided vol.	= na		(88 - 568)
residual vol.	= na		(5 - 42)
urine flow	= na		



B window %: UL=100 LL=C

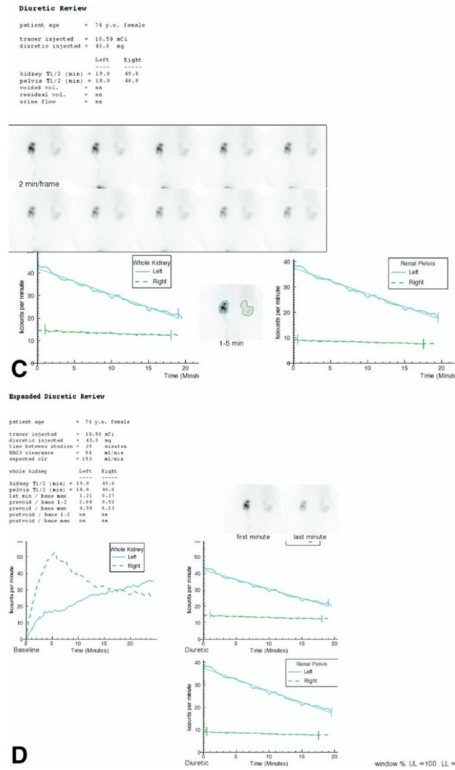


Figure 4. (A) The standard display shows demographic data, the dose injected, dose counted on the camera, percent dose infiltrated, the MAG3 clearance and the expected MAG3 clearance followed by the percent uptake, T_{max} , $T_{1/2}$, and 20-minute/max ratios for the whole kidney ROI. The voided volume, postvoid residual and urine flow rate were not measured. The upper central panel shows 2-second images at the beginning of the acquisition. The upper right panel shows the injection site; just beneath is a frame for viewing a dynamic cine, and pre and postvoid bladder images. The central panel shows twelve 2-minute images followed by a postvoid image of the kidneys with the patient lying on the camera in the same position as the initial images. The lower left panel shows the whole kidney ROIs and the whole kidney renogram curves; the lower right panel shows the cortical ROIs and the cortical renogram curves. The MAG3 clearance was reduced (94 mL/min/1.73 m²) compared with a normal range of 226–439 mL/min/1.73 m². The relative uptake of the left kidney was 26%. The $T_{1/2}$ of the left kidney was greater than 50 minutes and the $T_{1/2}$ of the right kidney was 19 minutes. The 20-minute/max ratio was bilaterally abnormal; consequently, the patient received furosemide followed by a second acquisition (Fig. 1C). (B) An expanded review is available to the reviewers. This display shows the patient values and normal ranges for the MAG3 clearance, residual urine volume, percent relative uptake and the T_{max} , 20-minute/max, $T_{1/2}$, and postvoid/max ratio for whole kidney and cortical ROIs. The expanded review page also shows an enlarged parenchymal image obtained at 2 to 3 minutes, an enlarged display of the 19- to 20-minute image, and quality control images showing the before and after injection syringe counts and time of the bolus arrival in the kidneys. (Color version of figure is available online.) (C) This panel shows displays the 2-minute sequential images after the administration of 43 mg of furosemide. The curves were generated from whole kidney and renal pelvic regions of interest. The $T_{1/2}$ of the left renal pelvis was 19 minutes and 46 minutes for the right renal pelvis. (D) This panel displays the baseline and furosemide acquisition on the same scale. The time activity curve generated by the pelvic

region of interest is also displayed on an expanded scale. Even though the right kidney is abnormal, tracer washed out of the renal pelvis and the ratio of kidney counts in the prevoid furosemide acquisition to the maximum counts was only 0.23. On a 5-point scale, obstructed, probably obstructed, equivocal, probably nonobstructed and nonobstructed, the experts interpreted the right kidney as probably nonobstructed. Activity washed out of the pelvis and the ratio of prevoid furosemide counts to the maximal counts on the baseline study was only 0.23. RENEX interpreted the right kidney as not obstructed (certainty factor of -0.42). Relative and absolute function of the left kidney were reduced, there was prominent pelvic retention, washout was prolonged with a pelvic $T_{1/2}$ of 19 minutes and the ratio of prevoid furosemide counts to the maximal counts on the baseline study was abnormal at 0.59. The consensus interpretation of the experts was probably obstructed; RENEX also interpreted the left kidney as obstructed (certainty factor of 0.34). (Color version of figure is available online.) interpretation of diuretic renal scans but also as an educational tool for students and trainees.

Table 1

Post-Void/Maximum Count Ratios Using Regions of Interest Over the Entire Kidney *

	N	Mean	SD	Minimum	5th Percentile	95th Percentile	Maximum
Post-void/max ratio right kidney	106	0.08	0.04	0.02	0.03	0.16	0.24
Post-void/max ratio left kidney *	106	0.09	0.03	0.03	0.05	0.15	0.20

SD, standard deviation.

Reprinted with permission from Esteves et al.⁴⁵

* There is a minor but significant difference in the post-void to maximum count ratios for the left kidney between younger (<40 years) and older (>40 years) adults. There is no significant difference between males and females.

Table 2

Residual Bladder Volume*

	Sex	Age	N	Mean	SD	Minimum	5th Percentile	95th Percentile	Maximum
Residual volume (mL)	M [‡]	40	12	9	7	0	0	23	23
	M [‡]	>40	16	30	28	8	8	91	91
	F	40	18 [‡]	15	10	5	5	36	36
	F	>40	16	17	9	5	5	42	42

SD, standard deviation.

Reprinted with permission from Esteves et al.⁴⁵

* There is no significant difference in residual bladder volume between males and females.

[‡] There is a significant difference in residual bladder volume between younger (<40 years) and older (>40 years) males.

[‡] One female patient had a residual volume of 256 mL; this value was considered to be abnormal and deleted from the analysis.

Table 3

Camera-Based MAG3 Clearances (mL/min/1.73 m²)

	Sex	N	Mean	SD	Minimum	5th Percentile	95th Percentile	Maximum
MAG3 clearance*	M	44	338	63	211	238	433	454
	F	62	309	71	188	226	439	503
All subjects		106	321	69	188	226	439	503

SD, standard deviation.

Reprinted with permission from Esteves et al.⁴⁵*The difference is significant ($P < 0.05$) between males and females.

Table 4

Most Frequent Kidney Parameters in the Training Set on Level 1

Kidney Parameter	Frequency (%)	
	Right Kidney	Left Kidney
Cortical 20-min/max ratio	32.1	3.6
Postvoid/max ratio	29.1	83.0
Postvoid/1- to 2-min ratio	13.3	6.3
19- to 20-min/max ratio	10.8	1.0

Reprinted with permission from Binongo et al.⁶⁴

Table 5

Comparison of RENEX and Expert Interpretations

Expert	RENEX		
	Obstructed	Equivocal	Non-obstructed
Obstructed	14	3	0
Equivocal	5	8	2
Non-obstructed	3*	9	73

* RENEX did not exclude obstruction based on the baseline acquisition. No furosemide was given; consequently, these kidneys were assigned a RENEX interpretation of obstruction.