



Published in final edited form as:

Semin Nucl Med. 2014 March ; 44(2): 146–158. doi:10.1053/j.semnuclmed.2013.10.007.

Computer assisted diagnosis in renal nuclear medicine: rationale, methodology and interpretative criteria for diuretic renography

Andrew T Taylor, MD and Ernest V Garcia, PhD

Department of Radiology and Imaging Sciences, Emory University School of Medicine

Abstract

The goal of artificial intelligence, expert systems, decision support systems and computer assisted diagnosis (CAD) in imaging is the development and implementation of software to assist in the detection and evaluation of abnormalities, to alert physicians to cognitive biases, to reduce intra and inter-observer variability and to facilitate the interpretation of studies at a faster rate and with a higher level of accuracy. These developments are needed to meet the challenges resulting from a rapid increase in the volume of diagnostic imaging studies coupled with a concurrent increase in the number and complexity of images in each patient data. The convergence of an expanding knowledge base and escalating time constraints increases the likelihood of physician errors. Errors are even more likely when physicians interpret low volume studies such as ^{99m}Tc -MAG3 diuretic scans where imagers may have had limited training or experience. Decision support systems include neural networks, case-based reasoning, expert systems and statistical systems. iRENEX (renal expert) is an expert system for diuretic renography that uses a set of rules obtained from human experts to analyze a knowledge base of both clinical parameters and quantitative parameters derived from the renogram. Initial studies have shown that the interpretations provided by iRENEX are comparable to the interpretations of a panel of experts. iRENEX provides immediate patient specific feedback at the time of scan interpretation, can be queried to provide the reasons for its conclusions and can be used as an educational tool to teach trainees to better interpret renal scans. iRENEX also has the capacity to populate a structured reporting module and generate a clear and concise impression based on the elements contained in the report; adherence to the procedural and data entry components of the structured reporting module assures and documents procedural competency. Finally, although the focus is CAD applied to diuretic renography, this review offers a window into the rationale, methodology and broader applications of computer assisted diagnosis in medical imaging.

© 2013 Elsevier Inc. All rights reserved.

Corresponding author: Andrew T Taylor, MD, Department of Radiology and Imaging Sciences, Emory University Hospital, 1364 Clifton Rd, NE, Atlanta, GA 30322, ataylor@emory.edu, Phone: 404-727-4855, Fax: 404-727-3488.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Introduction

Nuclear medicine physicians and radiologists have the responsibility to monitor scan performance, analyze sequences of images, assimilate clinical information, assess quantitative data, formulate an interpretation and generate a focused and coherent report. These processes are demanding and require knowledge of multiple disease processes as well as expertise in the performance and interpretation of a variety of imaging modalities; these processes can also be associated with multiple sources of error including a lack of procedural competency, cognitive bias, fatigue, insufficient training and lack of expertise.

Lack of training, experience or expertise—The problems of insufficient training, limited experience and lack of expertise simply reflect the inevitable convergence of a rapidly expanding knowledge base and escalating time constraints that characterize the training and practice of medicine; these problems are especially acute in the setting of low volume studies where training is even more likely to have been limited. For example, radiology residents receive only 4 months of training to cover all of nuclear medicine including nuclear cardiology yet radiologists perform the vast majority of non-cardiac nuclear medicine studies. In contrast, the American Board of Nuclear Medicine has determined that competency in nuclear medicine requires 36 months of training to master the same material radiology residents will attempt to assimilate in 4 months. Moreover, radiology residents are largely trained to interpret static or a series of static images that require a detailed knowledge of anatomy while interpretation of studies such as diuretic radionuclide renography depends much more on an understanding of renal physiology and ^{99m}Tc -MAG3 (mercaptoacetyltriglycine) pharmacokinetics. Even so, interpretation of low volume studies such as renal scintigraphy can be a challenge for fully trained nuclear medicine physicians since a large percentage of an estimated 590,000 renal scans are performed annually in the United States are interpreted at sites that perform fewer than 3 studies per week (1).

Because of differences in training and expertise, image interpretation is sometimes more dependent on which physician interprets the study rather than the presence or absence of underlying disease. Different specialists may disagree on the interpretation of the same study between 9% and 72% of the time (2). In a recent study, Hunsche showed that 25% of the diuretic renal scan interpretations (obstructed, equivocal, not obstructed) depended on which experienced physician interpreted the study (3). This problem is not limited to nuclear medicine but also applies to other areas of imaging such as MR imaging of the lumbar spine and low-dose CT screening (4, 5).

Cognitive bias—The two major components of clinical decision consist of making a diagnosis and devising a treatment plan based on that diagnosis; if the diagnosis is incorrect, there is a much greater likelihood that the treatment plan will also be incorrect or suboptimal. Although many factors contribute to diagnostic error, a principal cause consists of cognitive errors where the problem lies not in a lack of knowledge but with the clinician's thinking process (6). Our minds are quite vulnerable to cognitive biases, logical fallacies and false assumptions. Cognitive failures are best understood in the context of how our brains

manage and process information. The two principal modes can be categorized as System 1 and System 2 (7). System 1 operates automatically, intuitively and quickly, effortlessly originating impressions and feelings while System 2 requires mental effort, focused concentration and analysis. Cognitive errors are much more likely to occur when the scan, radiograph or clinical problem are analyzed by System 1. Examples of System 1 errors consist of reports that are internally contradictory or reports where diuretic renal scan interpretations are primarily based on a simple heuristic rule governing the T1/2 (If the T1/2 is greater than 20 min, the kidney is obstructed). The cognitive error rate may be even higher in areas where training, experience or expertise is compromised.

Lack of procedural competency—Many regulatory and oversight groups require peer review processes to evaluate the interpretative competence of radiologists and nuclear medicine physicians on an ongoing basis but systems for evaluating *procedural competency* are not widely available (8). Appropriate procedures are essential to a quality study and extensive procedure guidelines are provided by many professional organizations to facilitate procedural competency. To help standardize practice and guide interpretation of renal scans, for example, an international group of experts in renal nuclear medicine has published guidelines for renovascular hypertension (9), suspected obstruction (10), clearance measurements (11) and quality control (12); however, a recent survey showed that *less than 50% of full time* British nuclear medicine practitioners were even aware that a clearance guideline existed (13). The situation is undoubtedly worse in the US where the vast majority of the renal scans are interpreted by *part time* practitioners(1). Moreover, there has been a proliferation of guidelines; the National Guideline clearinghouse (<http://www.guideline.gov>) contains over 2373 clinical practice guidelines from 285 organizations. Too often physicians lack the time and/or motivation to locate, read, assimilate and implement the current and most appropriate guideline recommendations. Guidelines alone are not sufficient to ensure procedural competency; future research must focus on innovative intervention approaches that will help improve the translation of current knowledge into every day clinical practice (14). Decision support systems that function in the reading room can overcome the limitation of guidelines by providing immediate patient specific feedback at the time of scan interpretation and help to minimize interobserver variability by providing standardized, reproducible interpretative guidance.

Report generation—Too often, renal nuclear medicine reports fail to contain the essential elements required to evaluate and interpret the study, fail to document the technical components of the study necessary for accountability, quality assurance and reimbursement or fail to communicate the results to the referring physician in a clear and unambiguous structured report (15). Standardization of imaging reports represents an effort to counter this limitation.

Artificial Intelligence (AI)

Artificial intelligence has the capacity to address the problems of training, expertise, physician fatigue, cognitive bias, procedural competency and report generation and represents a domain of science dealing with the processes of modeling and problem solving. AI is a broad field and encompasses artificial neural networks (ANN), fuzzy logic, genetic

algorithms, case based reasoning, expert systems and predictive statistical systems reasoning (16-23). ANN are designed to model processing of information by neurons; genetic processes mimic the process of natural evolution while expert systems model human expertise in a given field of knowledge (24). Diagnostic imaging has now become largely digital and computers have become an essential component of acquiring and processing imaging studies thus facilitating the application of computer based artificial intelligence methods to broad areas of nuclear medicine including planar and SPECT myocardial perfusion studies SPECT (18, 25, 26), pulmonary embolism (27, 28), dementia (29), bone scan interpretation (30), detection of lung cancer (31), renal transplant evaluation (32), detection of renovascular hypertension (33, 34) and the evaluation of suspected renal obstruction (35, 36).

Case-based reasoning—Case-based reasoning involves the process of solving new problems based on the solutions of similar past problems. In imaging, an algorithm searches a library of patient cases to find the ones with features that best match the features of the patient study being analyzed and interprets the study using the cases with the best match.

Artificial Neural Networks—The artificial neural network (ANN) is an algorithm designed to model living neuronal function and can be trained to recognize similarities in patterns and learn by example. Neurons are grouped in layers and form connections with cells from each neighboring layer but lack connections within layers; one of the more straightforward types of artificial neural networks is the feed forward neural network in which the signal spreads in one chosen direction. A fully connected network of neural nodes arranged in input, hidden and output layers is shown in Figure 1.

Hamilton et al utilized a feed forward neural network to detect renal artery stenosis in renal transplants while Nielsen et al employed a feed forward neural network for the analysis of captopril augmented renal scintigraphy in suspected renovascular hypertension (33, 34). In these promising approaches, repeated recognition trials were run using sample ^{99m}Tc -mercaptoacetyltriglycine (MAG3) renal scan data as inputs and the corresponding renal angiography results or expert interpretation as outputs in order to optimize the strength between the input and output nodes and thus train the net so that the input data eventually predicts the output.

Predictive statistical system—Statistical methods for prediction incorporate scientific knowledge about the data as well as the variability of the measurements in the observed data. They also have the ability to make inferences from a sample to a population, test hypotheses and estimate of parameters of interest.

Expert (knowledge-based) systems—In these systems, a knowledge base of heuristic rules is obtained from human experts capturing how they make their interpretations. These heuristic rules usually take the form of “*If A, then B.*” expressions. Building on our previous expertise (25, 26), we have developed a generalized methodology to aid in the interpretation of imaging studies and have applied this generalized methodology to develop a comprehensive informed renal expert system (RENEX without clinical information, iRENEX with clinical information) for detecting renal obstruction using both clinical

information and data extracted from pre and post furosemide ^{99m}Tc -MAG3 acquisitions. The goal of the RENEX and iRENEX projects is to develop and implement a knowledge based decision support system that will integrate clinical information with automated image analysis to provide a comprehensive renal scan interpretation indistinguishable from that of experts.

Choosing an AI methodology

Two primary issues for any decision support system are its accuracy and its clinical acceptance. Both are necessary in terms of the ultimate clinical utility. Direct application of standard statistical methods such as logistic regression use well-understood mathematical techniques; however, standard statistical programs have three limitations 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 and they have no means for “explaining” (justifying their findings) to physicians (37); these limitations are also shared by ANN and case based reasoning.

While more sophisticated methods can overcome some of these deficiencies, knowledge-based systems offer a natural environment for implementing the tools that are lacking in many statistical, ANN and case based reasoning approaches and can provide the physician with a rationale for medical decisions (38). We chose to implement a knowledge based expert system for three reasons: (1) Expert systems can generally be developed with fewer numbers of cases than are required for statistical, ANN and case-based systems; (2) knowledge based systems have the capacity to justify or explain their conclusions; and (3) we had previous experience with this approach (26). In spite of these advantages, knowledge based systems may benefit from the insights derived from more formal mathematical approaches and we also chose to implement both expert system development and predictive statistical modeling with the expectation that each approach would inform and strengthen the other approach and the project would serve as a benchmark for the scientific comparison and collaboration of these two important fields of medical decision-making.

Which gold standard: expert interpretations or outcome?

Use of clinical outcome as the gold standard for a diuretic 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 (intervention versus observation); consequently, this gold standard is biased. An additional problem is illustrated by a patient who had a pyeloplasty to relieve obstruction one 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 non-obstructed

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 protocol for diuretic renography

Our protocol is based on the 1996 international consensus panel recommendations for diuresis renography (10). The consensus panel recommended the intravenous injection of ^{99m}Tc -MAG3 followed by a 35 minute continuous acquisition with furosemide administered at 20 minutes; an alternative protocol was to divide the continuous acquisition into two stages: a 20 minute baseline acquisition followed by furosemide administration and a second acquisition. Kuyvenhoven et al have pointed out that the inconvenience of furosemide administration can be omitted if the baseline scan can exclude obstruction (39). Our approach has been similar; we have used the two stage acquisition since 1990 and omitted Stage 2 when the baseline acquisition was normal and excluded obstruction. The details of our acquisition and processing protocols have been described in previous publications (35, 36, 40).

Normal ranges, reproducibility and renal size

Implementation of the expert system required clinical data that was either unavailable or insufficiently robust and led us to design studies to address areas of clinical uncertainty.

What are the ranges of normal values for the camera-based MAG3 clearance, renogram and voiding parameters?—Clearance measurements and other specific quantitative parameters can assist in scan interpretation and patient management (9-12, 41-47). However, to utilize these parameters in our decision support system, both normal as well as abnormal ranges had to be specified and assigned a confidence level. To determine normal ranges, archived ^{99m}Tc -MAG3 acquisitions from 106 subjects evaluated for kidney donation were processed using QuantEM™ 2.0, an in-house upgrade to the initial QuantEM™ software (48). The ^{99m}Tc -MAG3 clearance was obtained using a camera based clearance technique previously validated in a multicenter trial (49). This results showed the mean \pm sd for the BSA corrected camera-based ^{99m}Tc -MAG3 clearance in normal subjects to be 321 ± 69 mL/min/1.73 m², essentially identical to the mean \pm sd of the plasma sample ^{99m}Tc -MAG3 clearances of potential renal donors measured at different institutions, 304 ± 70 and 317 ± 74 mL/min/1.73 m² (50, 51). The percent relative uptake in the right and left kidneys was 49% and $51\% \pm 4\%$ respectively with no difference between males and females and the mean cortical 20 min/maximum count ratio was 0.19 with a SD of 0.07 and 0.04 for the right and left kidneys, respectively. The mean post-void/max whole kidney count ratio was < 0.1 and the mean post-void residual bladder volume was < 30 mL. These results confirmed and extended previous studies(52-54), established normal limits adjusted for age and gender (48) and provided the data to generate certainty factors (see below) for each of the parameters. Selected values are shown in Tables 1 and 2.

What is the normal variation in relative uptake in stable patients? How much of a change is required to be confident of a loss of function?—When patients have more than one renal study, rules have to be developed to evaluate changes in parameter values. For example, if the relative renal uptake of ^{99m}Tc -MAG3 changes from 50/50% to

55/45% in a patient with suspected obstruction, does the kidney with the 45% uptake require an intervention to preserve function or is that 5% change within the error of measurement? In spite of the importance of this question, there were no hard data defining the error in sequential ^{99m}Tc -MAG3 parameter measurements in stable patient populations. To address this question, we conducted a prospective study to measure the reproducibility of ^{99m}Tc -MAG3 parameter values and define the thresholds for change beyond that expected by chance (55). Using a perirenal background region of interest, measurements of relative renal uptake of ^{99m}Tc -MAG3 and common renogram parameters were highly reproducible; in regard to relative uptake, a decrease of 7% (i.e., 50% to 43%) implies a loss in renal function at the 95% confidence level (55).

Determined renal size based on the ^{99m}Tc -MAG3 images—Chronic renal diseases result in bilateral small kidneys whereas the kidneys may be bilaterally enlarged in early diabetic renal disease, acute interstitial nephritis, HIV nephropathy, and amyloidosis (56-60). A unilateral small kidney is obvious but a symmetrical bilateral increase or decrease in renal size is not likely to be recognized on a ^{99m}Tc -MAG3 scan. Correlative imaging studies may provide size information but unpublished results from our institution indicate that approximately 50% of patients referred for renal scintigraphy lack prior imaging studies and even when imaging reports are available, the report often fails to comment on renal size. To be able to use renal size to assist in scan interpretation, we had to determine normal ranges. Again, our study population consisted of 106 normal adults evaluated for kidney donation who underwent ^{99m}Tc -MAG3 renal scintigraphy. Renal length, width and area were determined from the maximum pixel length, pixel width and area of whole kidney ROIs and correlated with patient gender, height, weight, body mass index and body surface area and reference values generated using a quantile regression model (61, 62). Regression equations defining the upper and lower limits of area and length are shown in Table 3.

Techniques to minimize processing errors and reduce intra- and interobserver variability

Processing errors can lead to mistakes in interpretation by both physicians and AI systems. Moreover, there is a need to reduce intra- and interobserver variability to increase the likelihood that the same quantitative data will be generated regardless of where a study is performed or which technologist or physician processes the data.

Quality control software to monitor technologist input—We have developed software to check the entire technologist input data used by RENEX (63). Checks are made for logical inconsistency (negative and non-numeric 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 tracer injection) as well as factors that may lead to unreliable quantitative values. Prospective evaluation of this quality control software demonstrated that QC issues are not uncommon and that they can be identified and often corrected. Even if QC issues are recognized but cannot be corrected, they can be flagged for the reviewing physician and certainty factor (see below) associated

with the suspect parameter can be modified in the expert system so that the questionable values have less effect on the final interpretation.

Automatic assignment of kidney ROIs—A robust image-processing algorithm (AUTOROI) to automatically detect whole-kidney contours and generate renal regions of interest (ROI) has been developed for the extraction of quantitative data thus providing an objective and reproducible approach to ROI generation (64).

Detection and correction of errors due to patient motion—Patient motion during the 20-30 min dynamic acquisition can partially shift the kidney outside the assigned ROI causing errors in the calculated renal function parameters which can lead to errors in interpretation. To address this problem, we developed an algorithm to detect and correct patient motion and then simulated 10 types of vertical motion in 86 patient studies, resulting in 860 image sets (65). The algorithm accurately detected motion as small as 0.25 pixel and detected and corrected whole-pixel motion with a high degree of accuracy. Importantly, by accurately identifying unshifted frames where no motion is simulated, the algorithm helps to prevent the introduction of errors during motion correction (65).

The Architecture of iRENEX

The architecture of iRENEX is summarized in Fig 2 and discussed in greater detail below:

Certainty Factors—RENEX and iRENEX were inspired by two previously developed expert systems (25, 26, 38). In these systems, each quantitative parameter must be converted into a certainty factor to express the degree of confidence regarding the normality or abnormality of that parameter. A certainty factor value of + 1.0 was assigned to indicate that a parameter is “definitely abnormal” while a certainty factor value of - 1.0 is assigned to indicate that a parameter is “definitely normal.” The certainty factor to indicate the boundary between equivocal and abnormal is assigned a value of + 0.2, a truly equivocal value is assigned a CF of 0 and the certainty factor value of -0.2 was assigned to indicate the boundary between equivocal and normal. A certainty factor increasing from +0.2 to +1.0 indicates increasing confidence in that the parameter is abnormal with the value “+1.0” representing 100% confidence.

To develop certainty factors for parameters extracted from the renogram, an expert in radionuclide scintigraphy (domain expert) combined his experience and the normal limits for the kidney parameters extracted from the Tc-99m MAG3 scans of 106 potential renal donors (48) 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 convert the value of any individual quantitative parameter to a certainty factor (Fig 3) The collection of these sigmoid curves for all the variables constitutes the parameter knowledge library

The Knowledge Base and Inference Engine—Sixty heuristic rules (*If A, then B*) were initially extracted from the domain expert to generate the knowledge base for detecting obstruction; 12 of these 60 rules were specifically applied to the baseline study to determine the need for furosemide administration while the remainder were applied to the diuretic

acquisition or to relationships between parameters of the baseline and diuretic acquisitions to evaluate the presence or absence of obstruction.

An inference engine is software that selects and executes the rules to reach a conclusion. The RENEX inference engine approximates Bayes theorem to combine the certainty factors, known as Bayesian combinatorics, to reach a conclusion which is expressed by a combined certainty factor (38). One of the rules, for example, states, “If the ratio of the post-void kidney counts of the post-furosemide 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 not obstructed.”

When the inference engine begins to analyze a study, 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. 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 in the conclusion of obstruction. If the certainty factor is less than -0.2, the kidney is not obstructed, and if it lies between -.2 and +.2, the kidney is equivocal for obstruction. The initial 60 rules were modified as the system was trained with patient data and additional rules were added.

The Justification Engine: Do RENEX and iRENEX get the right answer for the right reason?—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 in order to track and justify the logic of the conclusions (66). 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 presence or absence of obstruction (66, 67).

Capturing and quantifying clinical information

Capturing clinical information—Our goal was to capture all the clinical information that might be important to clinicians for the interpretation of the study such as the serum creatinine, a clinical finding of left flank pain following furosemide administration or data extracted from imaging reports (type of study, date and findings). FileMaker Pro was used to develop a relational database that could (1) capture clinical information, (2) link this clinical information to the patient file and the quantitative renal scan parameters, (3) export the data to RENEX, and (4) export a summary report as a text file for physicians. This database was chosen because of its widespread commercial use, flexible review layouts and the capacity for scripting (the development of customized programs to enhance functionality). The database contained 342 fields for each patient to accommodate the range of clinical information that might potentially be available and included summaries of CT, US, MR, IVP, prior MAG3 scans, surgeries, stents and serum creatinine values (68). The information that the clinician considers to be relevant is extracted from the medical records and entered

into the data base. This process mirrors the current retrieval of supplemental information by the clinician; in the future, medical records may be amenable to data mining.

Conversion of non-quantitative variables (e.g, results of imaging studies) into certainty factors—The inference engine for RENEX and iRENEX requires a certainty factor (CF) for each quantitative and non-quantitative variable. The process of developing CFs for non-quantitative variables is illustrated using hydronephrosis. An imaging report, for example, might describe hydronephrosis as present or absent and, if present, might further characterize the hydronephrosis as mild, moderate or severe. Each of these features was initially assigned a certainty factor by the domain expert to express the certainty that the abnormality was present. If hydronephrosis was described as present but not qualified by a descriptive adjective, the assigned CF was 0.35; mild, moderate and severe hydronephrosis were assigned CFs of 0.25, 0.45 or 0.8, respectively; if the image included the kidneys in the field of view but the report had no comment regarding hydronephrosis, hydronephrosis was assumed to be absent with a CF of - 0.4. These initial CFs were then modified by experience and the IKLA algorithm described below to reach the same conclusion regarding the presence or absence of obstruction as expert readers (69, 70)

What is the impact of clinical information on scan interpretation?

To facilitate incorporation of clinical information into the expert system and to help guide clinicians, a study was conducted to determine the impact of clinical information on diuretic renography scan interpretation. Baseline and furosemide MAG3 acquisitions of 92 patients with suspected obstruction were randomly selected from an archived database and independently interpreted by 3 experienced readers without clinical information and re-interpreted with access to a File Maker Pro database. All studies included a camera-based MAG3 clearance. Readers scored each kidney on a scale from +1.0 to -1.0; scores > 0.21 represented obstruction with higher scores indicating greater confidence in the diagnosis of obstruction. Scores from +0.2 to -0.2 were equivocal and scores < -0.2 indicated non-obstructed kidneys. There was a subsequent consensus reading without and then with clinical information where the 3 readers were allowed to change their scores; the mean score served as the final score.

When the initial scan interpretation was non-obstructed, clinical history was essentially irrelevant. Among the scans initially interpreted as obstructed or equivocal, clinical history changed the diagnostic category in only 2-3% of kidneys. Moreover, the impact of clinical history varied among readers with most important clinical factors affecting interpretation being the presence of a ureteral stent and the presence or absence of hydronephrosis or hydroureter (71). To provide a more objective basis for the use of clinical history elements in scan interpretation, this study has been expanded and is the subject of a manuscript recently submitted for publication. iRENEX has the advantage of providing an objective, reproducible approach to evaluate and standardize implementation of clinical history into scan interpretation and can help determine the utility of different clinical parameters.

Further selection and weighting of variables for scan interpretation

Two additional approaches were used to supplement input from the domain expert and data from clinical studies to determine the key variables, select the certainty factors and appropriately weight the rules in the RENEX knowledge base.

Key variables based on a statistical predictive model—Logistic regression analysis and proportional odds modeling were successfully utilized to develop a predictive system for the interpretation of ^{99m}Tc -MAG3 diuretic scans and to independently identify key interpretative variables impacting scan interpretation (72). Common variables included in the analysis were the time to maximum counts, $T_{1/2}$ and the 20 min/maximum count ratio for both whole kidney and cortical regions of interest, relative uptake, pelvic $T_{1/2}$ and the ratio of post-void kidney counts to the 1-2 min and maximum kidney counts (72). The single most important baseline variable for excluding obstruction on the baseline examination was the ratio of post-void counts to maximum counts while the single most important variable in evaluating obstruction following furosemide was the ratio of renal counts in the last minute of the furosemide acquisition to the maximum counts on the baseline examination (the ratio of the ratio post-furosemide post-void kidney counts to the maximum baseline kidney counts would likely have had an even greater predictive power but post-furosemide post-void kidney counts were not recorded in enough patients for us to perform this analysis). The complete results have been described in detail (72).

Key variables based on an international panel of expert readers—A panel from the International Scientific Committee of Radionuclides in Nephro-urology (ISCORN; <http://www.iscorn.org>) developed a Guidance Document for quality assurance and structured reporting of diuresis renography in adults (15). This document includes a listing and discussion of elements that are considered to be essential or important for scan interpretation and quality control. All elements recommend by the panelists (15) have been incorporated into RENEX.

How can the thousands of possible combinations of rules and certainty factors be optimized?

The continuing development of iRENEX has resulted in a major expansion of the RENEX knowledge base requiring new rules to incorporate (1) advanced quantitative features, (2) quality control data (3) clinical data, (4) knowledge mined through use of the justification engine and (5) determination of the support and confidence of each rule in the knowledge base. The vast combination of possible rules and certainty (weighting) factors per rule exceeds the capacity of any individual to optimize diagnostic accuracy by tracking the effect of multiple changes in rules and CFs on accuracy through an extensive patient data base. To address this problem, we developed and validated a novel accuracy-guided iterative knowledge-learning algorithm (IKLA) to optimize diagnostic accuracy (70).

This algorithm consists of implementing a) batch mode interpretation of patient studies, b) batched iterations of different knowledge islands defined as follows: 1) MAG3 quantitative rules for detecting obstruction, 2) quantitative rules for detecting the absence of obstruction, 3) renal clearance rules, 4) diuretic dose rules, 5) clinical rules, 6) QC rules, 7) rules/CFs

that individually optimize left and right kidneys 8) rules/CFs obtained from support/confidence analysis. IKLA iterated through the 256 (2^8) possible permutations of knowledge islands on a pilot group of 100 kidneys and a prospective group of 110 kidneys and the algorithm automatically determined the accuracy for each iteration. The accuracy of IKLA in the pilot group predicted the accuracy in the prospective group. Moreover, the highest accuracy was obtained for the entire set of rules/CFs demonstrating that each set contributed to overall accuracy (70). This approach has the potential to assist in the optimization of other complex decision support systems and has been submitted for publication.

RENEX agrees with expert interpretations

A pilot study demonstrated that RENEX could be optimized to perform comparably to experts (35). Following the pilot study, RENEX was prospectively evaluated in 95 randomly selected furosemide augmented patient studies (185 kidneys) obtained for suspected obstruction (36). Three experts, blinded to clinical information, independently graded each kidney as obstructed/probably obstructed, equivocal, and probably non-obstructed/non-obstructed; experts resolved differences by a consensus reading. These 3 expert categories were compared to the obstructed, equivocal and non-obstructed interpretations provided by RENEX. Agreement was assessed using weighted kappa and the predictive accuracy of RENEX compared to expert readers was assessed by the area under receiver operating curves (ROC).

RENEX showed good agreement with expert interpretations; specifically, RENEX agreed with the consensus reading in 84% (101/120) of non-obstructed kidneys, in 92% (33/36) of obstructed kidneys and 45% (13/29) of equivocal kidneys (36). The weighted kappa between RENEX and the consensus reading was 0.72 and was comparable to the weighted kappa between experts. There was no significant difference in the areas under the ROC curves when RENEX was compared to each expert using the other two experts as the gold standard (36).

The performance of RENEX is *equivalent* to an expert

Common statistical approaches to evaluate accuracy include ROC and kappa analysis but both methods have significant limitations and cannot answer the question of *equivalence*. Using log-linear modeling to analyze the 95 randomly selected furosemide augmented patient studies described above, RENEX was demonstrated to be *equivalent* to any expert; this equivalence was particularly strong in the obstructed and non-obstructed categories (73).

iRENEX: Incorporation of Clinical Information

Fifty-six new clinical rules regarding the presence or absence of obstruction were extracted from the domain experts and added to the 60 rules for RENEX. All the clinical rules were implemented after iRENEX reached a conclusion regarding obstruction based on the quantitative ^{99m}Tc -MAG3 parameters. The entire system was fine tuned and tested using a pilot group of 51 patients deemed by an expert panel to have 61 unobstructed kidneys and 38 obstructed findings (69). Processing time per patient was practically instantaneous using a 3.0 GHz PC programmed using IDL. When iRENEX and the experts used only

quantitative ^{99m}Tc -MAG3 data, RENEX agreed with the expert panel in 87% (34/39) of the obstructed kidneys and in 90% (55/61) of the unobstructed kidneys. Using a clinical database developed from patient histories and imaging reports, iRENEX agreed with the expert panel in 95% (37/39) of obstructed and 92% (56/61) non-obstructed kidneys. The clinical information significantly ($p < 0.001$) increased iRENEX certainty in detecting obstruction over using the quantitative data alone. In addition, there was a high correlation ($R = 0.92$) between the certainty factors (CF) of iRENEX and the median CF of the three experts regarding the presence or absence of obstruction demonstrating that iRENEX exhibited the same confidence level as experts in regard to the presence or absence of obstruction (Fig 4). These encouraging results warrant a prospective study in a large population of patients with and without renal obstruction to establish the diagnostic performance of iRENEX.

Performance of iRENEX Compared to Nuclear Medicine Residents

Three expert readers, iRENEX and 3 nuclear medicine residents with 10-12 months of training independently interpreted 41 randomly selected diuretic renography studies without clinical information, then reinterpreted the studies with clinical information provided in text file format from the relational database (74). Interpretations were scored on a continuous scale and ranged from +1.0 (definitely obstructed) to -1.0 (definitely not obstructed); the range from +0.2 to -0.2 was indeterminate. Without clinical information, the overall concordance correlation coefficient (CCC) between iRENEX and the experts was 0.890, significantly higher than the overall CCC between the residents and the experts, 0.688, $p < 0.05$ (Fig 5). With clinical information, the overall CCC between iRENEX and the experts was 0.876, also significantly higher than the overall concordance between the residents and the experts, 0.673, $p < 0.05$ (Fig 5). iRENEX agreed with the experts better than the residents and thus has the potential to serve as a teaching tool to help residents learn to interpret studies at the level of experts. To gain additional data, the comparisons between residents, experts and iRENEX have been substantially expanded and are the subject of a manuscript recently submitted for publication.

Acceptance of Computer aided diagnosis (CAD) in medicine and imaging

The application AI to nuclear medicine, radiology and other areas of medicine sometimes leads to a certain uneasiness of the part of the clinician. As an expression of this uneasiness, an anonymous reviewer of one of our papers wrote, "I would rather have a good technologist than a good computer program (as if the two were incompatible)," while another asked, "If the computer interprets the study, how is the radiologist going to get paid?" Underlying these questions is the fear that the imager may eventually follow in the footsteps of the fireman on a coal powered locomotive, no longer needed when the engine switches to diesel.

This fear is misplaced. The goal of artificial intelligence, expert systems, decision support systems and computer assisted diagnosis in imaging is the development and implementation of software to assist in the detection and evaluation of abnormalities, to alert the physician to cognitive biases, to reduce intra and interobserver variability and to allow physicians to interpret studies at a faster rate and with a higher level of accuracy. In support of these goals,

the Second Annual American College of Radiology Forum has encouraged the adoption of innovative solutions such as computer aided detection and diagnosis of disease (75). This has been our focus. Although our immediate goal has been the analysis and interpretation of diuretic renograms, components of our approach apply broadly to image analysis and include: (1) Extraction of quantitative parameters from diagnostic images related to a disease process; (2) Conversion of these quantitative parameters to certainty factors; (3) Conversion of descriptive parameters extracted from patient demographics and other test results to certainty factors; (4) Use of these quantitative and demographic parameters as input to rules in a knowledge base (production rules) extracted from domain experts; (5) Use of Bayesian combinatorics to infer conclusions, i.e. diagnostic findings; (6) Tracking the rules used to reach a conclusion and to justify that conclusion; (7) Translation of the conclusions to a structured textual report; and (8) Propagation of the parameters and conclusions to create a patient database which may be queried. These generic components can be adapted to facilitate the development of other targeted decision support systems in medicine and radiology. In summary, CAD has the potential to reduce the burdens of illness and disability by lowering costs, improving quality and increasing diagnostic accuracy.

Acknowledgments

This work was supported by a grant from the National Institute of Biomedical Imaging and Bioengineering (NIBIB) and the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK): R01 EB008838.

References

1. IMV Medical Information Division. 2003 nuclear medicine census market summary report. Des Plaines, IL: 2003. p. IV
2. Jaksic E, Beatovic S, Paunkovic N, et al. Variability in interpretation of static renal scintigraphy findings. *Vojnosanitetski Pregled*. 2005; 62:189–93.
3. Hunsche, AA. Porto Alegre, Rio Grande do Sul, Brazil: Federal University of Rio Grande do Sul; 2006. Value of quantitative data in the interpretation of diuresis renography for suspected urinary tract obstruction.
4. Carrino JA, Lurie JD, Tosteson ANA, et al. Lumbar spine: reliability of MR imaging findings. *Radiol*. 2009; 250:161–70.
5. Gierada DS, Pilgram TK, Ford M, et al. Lung cancer: interobserver agreement on interpretation of pulmonary findings at low-dose CT screening. *Radiol*. 2008; 246:265–72.
6. Croskerry P. From mindless to mindful practice – cognitive bias and clinical decision making. *N Engl J Med*. 2013; 368:2445–2448. [PubMed: 23802513]
7. Kahneman, D. 19-30 in *Thinking Fast and Slow* Ultrasonography and nuclear medicine. In: Schrier, RW.; Coffman, TM.; Falk, RJ.; Molitoris, BA.; Neilson, EG., editors. *Schrier's Diseases of the Kidney*. 9th. Lippincott, Williams and Wilkins; Philadelphia, PA: 2013. p. 346-376.
8. Mendiratta-Lala M, Eisenberg RL, Steele JR, Boiselle PM, Kruskal JB. Quality initiatives: measuring and managing the procedural competency of radiologists. *Radiographics*. 2011; 31(5): 1477–88. [PubMed: 21719719]
9. 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]
10. O'Reilly P, Aurell M, Britton K, Kletter K, Rosenthal L, Testa T. Consensus on diuresis renography for investigating the dilated upper urinary tract. *J Nucl Med*. 1996; 37:1872–1876. [PubMed: 8917195]
11. Blafox 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]

12. 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.
13. Cosgriff, PS.; Stevens, D. Impact of radionuclides in Nephrourology guidelines. www.alasbimjournal.cl
14. Hart PD, Bakris GL. Hypertension control rates: time for translation of guidelines into clinical practice. *Am J Emnd.* 2004; 117:62–64.
15. Taylor AT, Blaufox MD, De Palma D, et al. Guidance document for structured reporting of diuresis renography. *Semin Nucl Med.* 2012; 42:41–48. [PubMed: 22117812]
16. Haddad M, Adlassnig KP, Porenta G. Feasibility analysis of a case-based reasoning system for automated detection of coronary heart disease from myocardial scintigrams. *Artificial Intelligence in Medicine.* 1997; 9:61–78. [PubMed: 9021059]
17. Fujita H, Katafuchi T, Uehara T, Nishimura T. 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]
18. 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]
19. Hamilton D, Riley PJ, Miola UJ, Amro AA. A feed forward neural network for classification of bull's-eye myocardial perfusion images. *Eur J Nucl Med.* 1995; 22:108–115. [PubMed: 7758496]
20. 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]
21. Lindahl D, Palmer J, Pettersson J, White T, Lundin A, Edenbrandt L. Scintigraphic diagnosis of coronary artery disease: myocardial bull's-eye images contain the important information. *Clin Physiol.* 1998; 18(6):554–561. [PubMed: 9818161]
22. Lindahl D, Lanke J, Lundin A, Palmer J, Edenbrandt L. Improved classifications of myocardial bull's-eye scintigrams with computer-based decision support system. *J Nucl Med.* 1999; 40(1):96–101. [PubMed: 9935064]
23. Lindahl D, Palmer J, Edenbrandt L. Myocardial SPET: artificial neural networks describe extent and severity of perfusion defects. *Clin Physiol.* 1999; 19(6):497–503. [PubMed: 10583343]
24. Swietlik D, Bandurski T, Lass P. Artificial neural networks in nuclear medicine. *Nucl Med Review.* 2004; 7:58–67.
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. 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]
28. Fisher RE, Scott JA, Palmer EL. Neural networks in ventilation-perfusion imaging. Part 1. Effects of interpretative criteria and network architecture. *Radiol.* 1996; 198:699–706.
29. Mosconi L, Tsui WH, Herholz K, et al. Multicenter standardized 18F-FDG PET diagnosis of mild cognitive impairment, Alzheimer's disease and other dementias. *J Nucl Med.* 2008; 49:39–398. [PubMed: 18077521]
30. Sadik M, Jakobsson D, Ologsson F, et al. A new computer-based decision-support system for the interpretation of bone scans. *Nucl Med Commun.* 2006; 27:417–423. [PubMed: 16609352]
31. Gutte H, Jakobsson D, Olofsson F, et al. Automated interpretation of PET/CT images in patients with lung cancer. *Nucl Med Commun.* 2007; 28:79–84. [PubMed: 17198346]
32. Abdolmaleki P, Movhead M, Taniguchi RI, et al. Evaluation of complications of kidney transplantation using artificial neural networks. *Nucl Med Commun.* 1997; 18:623–630. [PubMed: 9342099]

33. 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]
34. Nielsen M, Granerus G, Ohlsson M, et al. Interpretation of captopril renography using artificial neural networks. *Clin Physiol Funct Imaging.* 2005; 5:293–296. [PubMed: 16117733]
35. 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]
36. Taylor A, Garcia EV, Binongo J, et al. Diagnostic performance of an expert system for interpretation of Tc-99m MAG3 scans in suspected renal obstruction. *J Nucl Med.* 2008; 49:216–224. [PubMed: 18199609]
37. Gorry G. Computer-assisted clinical decision making. *Met Info Med.* 1973; 12:45–51.
38. Shortliffe, EH. *Computer-Based Medical Consultations: MYCIN.* Elsevier scientific publishing company; Amsterdam, Netherlands: 1976. p. 264
39. Kuyvenhoven J, Piepsz M, Ham H. When could the administration of furosemide be avoided? *Clin Nucl Med.* 2003; 28:732–7. [PubMed: 12972994]
40. Taylor A, Manatunga A, Garcia EV. Decision support systems in diuresis renography. *Sem Nucl Med.* 2008; 38:67–81.
41. 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.
42. Taylor, A., Jr; Blaufox, MD.; Dubovsky, EV., et al. Procedure guideline for the diagnosis of renovascular hypertension. www.snm.org/guidelines
43. Li Y, Russell CD, Palmer-Lawrence J, Dubovsky EV. Quantitation of renal parenchymal retention of technetium-99m-MAG3 in renal transplants. *J Nucl Med.* 1994; 35:846–850. [PubMed: 8176469]
44. Piepsz A, Tondeur M, Ham H. NORA: a simple and reliable parameter for estimating renal output with or without frusemide challenge. *Nucl Med Commun.* 2000; 21:317–323. [PubMed: 10845219]
45. 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]
46. 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]
47. 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]
48. Esteves FP, Taylor A, Manatunga A, Folks RD, Krishnan M, Garcia EV. 99mTc-MAG3 renography: normal values for MAG3 clearance and curve parameters, excretory parameters, and residual urine volume. *AJR Am J Roentgenol.* 2006; 187:W610–617. [PubMed: 17114514]
49. Taylor A Jr, 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]
50. 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]
51. El-Galley R, Clarke HS, O'Brien DP, Taylor A. Normal parameters for Tc-99m MAG3 renography. *J Nucl Med.* 1998; 39:87P.
52. Klingensmith WC 3rd, 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.* Oct; 1994 35(10):1612–1617. [PubMed: 7931658]
53. 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(1):11–16. [PubMed: 9519415]
54. 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.
55. Taylor A, Manatunga A, Halkar R, Issa MM, Shenvi NV. A 7% decrease in the differential uptake of MAG3 implies a loss of renal function. *Urology.* 2009; 76:1512–1516. [PubMed: 20708778]

56. Segal MC, Lecky JW, Slasky RS. Diabetes mellitus: The predominant cause of bilateral renal enlargement. *Radiology*. 1984; 153:341–2. [PubMed: 6484164]
57. Rodriguez-de-Velasquez A, Yoder IC, Velasquez PA, et al. Imaging the effects of diabetes on the genitourinary system. *RadioGraphics*. 1995; 15:1051–1068. [PubMed: 7501850]
58. Rahbari-Oskoui, F.; Taylor, AT.; O'Neill, WC. Ultrasonography and nuclear medicine. In: Schrier, RW.; Coffman, TM.; Falk, RJ.; Molitoris, BA.; Neilson, EG., editors. *Schrier's Diseases of the Kidney*. 9th. Lippincott, Williams and Wilkins; Philadelphia, PA: 2013. p. 346-376.
59. Di Fiori JL, Rodrigue D, Kaptein EM, et al. Diagnostic sonography of HIV-associated nephropathy: New observations and clinical correlation. *AJR*. 1998; 171:713–16. [PubMed: 9725302]
60. Schutz K, Siffing PA, Forrest TS, et al. Serial renal sonographic changes in preeclampsia. *J Ultrasound Med*. 1990; 9:415–18. [PubMed: 2197424]
61. Koenker R, Bassett G. Regression quantiles. *Econometrica*. 1978; 46:33–50.
62. Taylor AT, Shenvi N, Folks RD, et al. Reference values for renal size obtained from MAG3 scintigraphy. *Clin Nucl Med*. 2012; 38:13–17. [PubMed: 23235485]
63. 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*. 2007; 35:27–33.
64. Garcia E, Folks R, Pak S, Taylor A. Totally automatic definition of renal regions-of-interest from Tc-99m MAG3 renograms: Validation in patients with normal kidneys and in patients with suspected renal obstruction. *Nucl Med Commun*. 2010; 31:366–374. [PubMed: 20145580]
65. Folks RD, Manatunga D, Garcia EV, Taylor AT. Automated patient motion detection and correction in dynamic renal scintigraphy. *J Nucl Med Tech*. 2011; 39:131–139.
66. Garcia EV, Taylor A, Manatunga D, Folks R. A software engine to justify the conclusions of an expert system for detecting renal obstruction on 99mTc-MAG3 scans. *J Nucl Med*. 2007; 48:463–470. [PubMed: 17332625]
67. Porenta G. Being right for the right reason: better than just being right? *J Nucl Med*. 2007; 48:335–336. [PubMed: 17332608]
68. Folks RD, Savir-Baruch B, Garcia EV, Verdes L, Taylor AT. Development of a relational database to capture and merge clinical history with the quantitative results of radionuclide renography. *J Nucl Med Technol*. 2012; 40:236–243. [PubMed: 23015477]
69. Garcia EV, Taylor A, Folks R, Manatunga D, Halkar R, Savir-Baruch B, Dubovsky E. iRENEX: A clinically-informed decision support system for the interpretation of Tc-99m MAG3 scans to detect renal obstruction. *Eur J Nucl Med Mol Imag*. 2012; 39:1483–1491.
70. Folks RD, Garcia EV, Manatunga D, Taylor AT. Novel accuracy-guided iterative knowledge-learning algorithm to optimize detection of renal obstruction from Tc-99m MAG3 renography. *J Nucl Med*. 2013; 54(5 (Suppl. 2)):508P.
71. Taylor AT, Savir-Baruch B, Folks RD, Halkar RK, Dubovsky EV, Garcia EV. Impact of clinical information on diuretic scan interpretation. *J Nucl Med*. 2010; 51:118P–119P.
72. Bao J, Manatunga A, Binongo JNG, Taylor A. Key Variables for interpreting MAG3 diuretic scans: development and validation of a predictive model. *AJR*. 2011; 197:325–333. [PubMed: 21785077]
73. Manatunga A, Binongo JNG, Taylor AT. Computer-aided diagnosis of renal obstruction: utility of log-linear modeling versus standard ROC and kappa analysis. *EJNMMI Research*. 2011; 1:5.
74. Taylor AT, Shenvi NV, Folks RD, et al. Tc-99m MAG3 diuretic renography: Impact of a decision support system (iRENEX) on resident interpretations. *J Nucl Med*. 2013; 54:171P.
75. Hillman BJ, Neiman HL. *Radiology 2012: Radiology and radiologists a decade hence – A strategic analysis for radiology from the Second Annual American College of Radiology FORUM*. *Radiol*. 2003; 227:9–14.

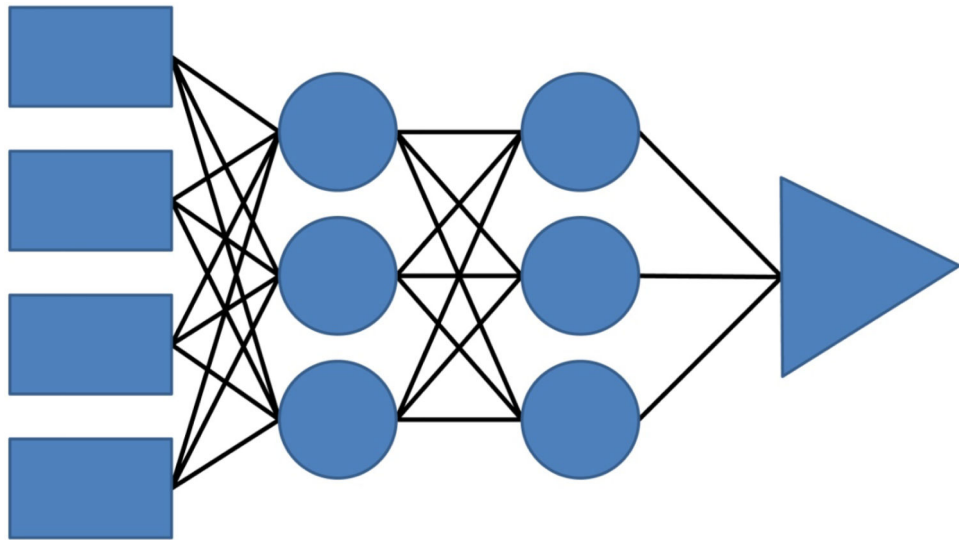


Fig 1.

A four layer feed forward (left to right) artificial neural network showing the input nodes as rectangles, the hidden nodes as circles, the output node as a triangle and connecting weights as lines.

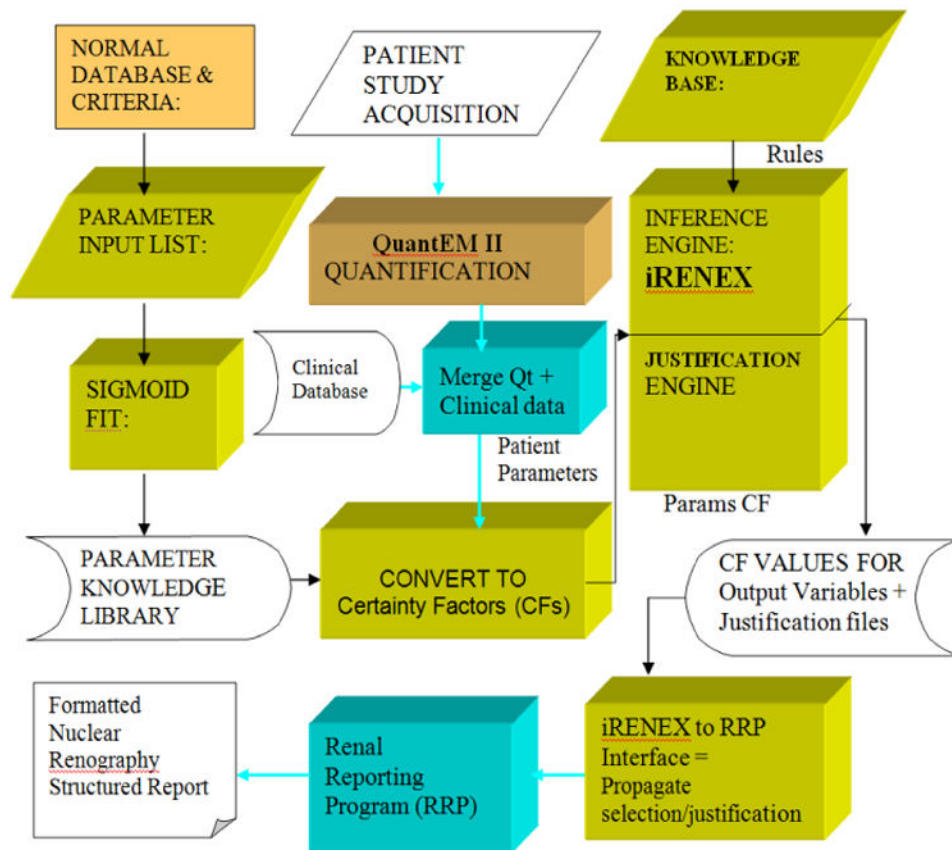


Fig 2.

Flow diagram for iRENEX. This diagram shows the flow depicting how a renal scan is acquired, processed, and quantified to extract parameters of renal function and how these parameters are converted to certainty factors that are then input to the expert system. iRENEX contains a library to translate parameter values to certainty factors, a knowledge base with 116 heuristic interpretation rules, a forward chaining inference engine to determine obstruction and a justification engine which records the sequence of each rule that is fired and the current certainty factor value of all input and output parameters at the time of instantiation in order to track and justify the logic of the conclusions. The trapezoidal blocks indicate domain expert; the rectangular blocks indicate software algorithms.

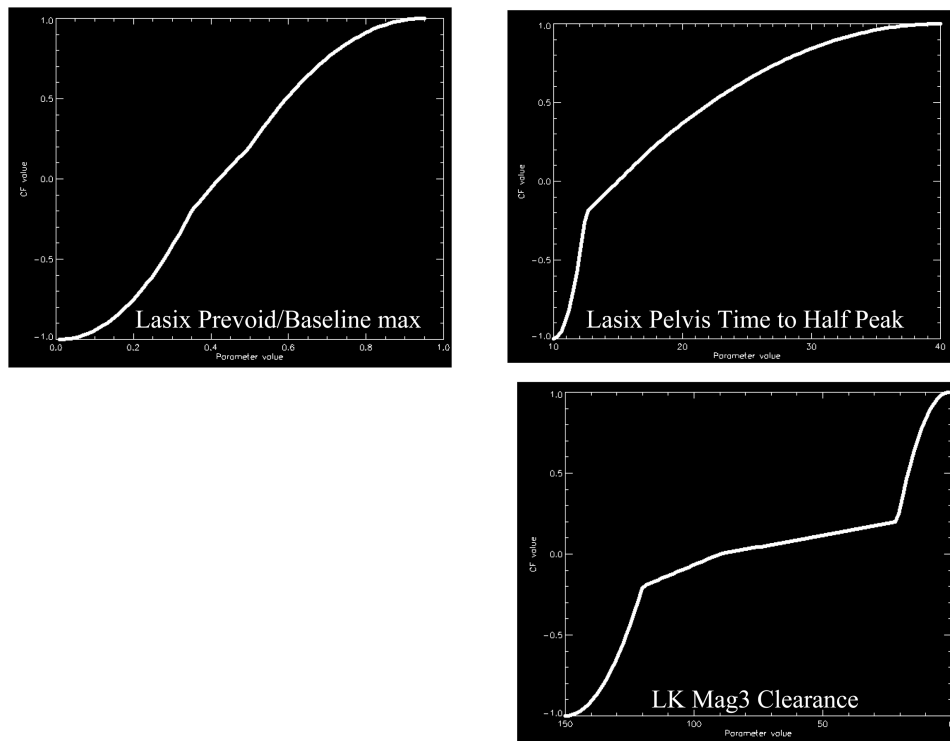


Fig 3.

Graphical representation of the transformation of 3 input quantitative parameters to certainty factor values. The 3 parameters are illustrated as follows. A: Lasix pre-void to baseline max (the ratio of the counts in the kidney ROI during the last frame of the post-furosemide renogram to the maximum counts in the kidney ROI from the pre-furosemide 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 (40 reproduced by permission).

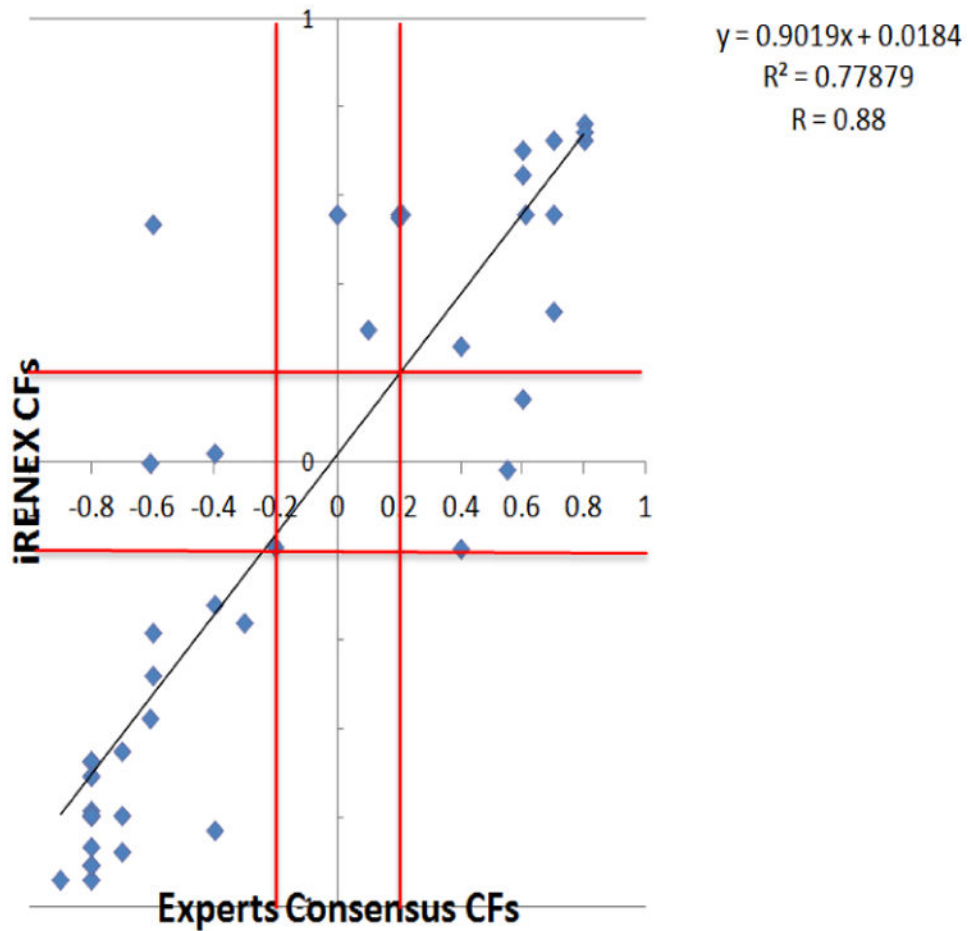
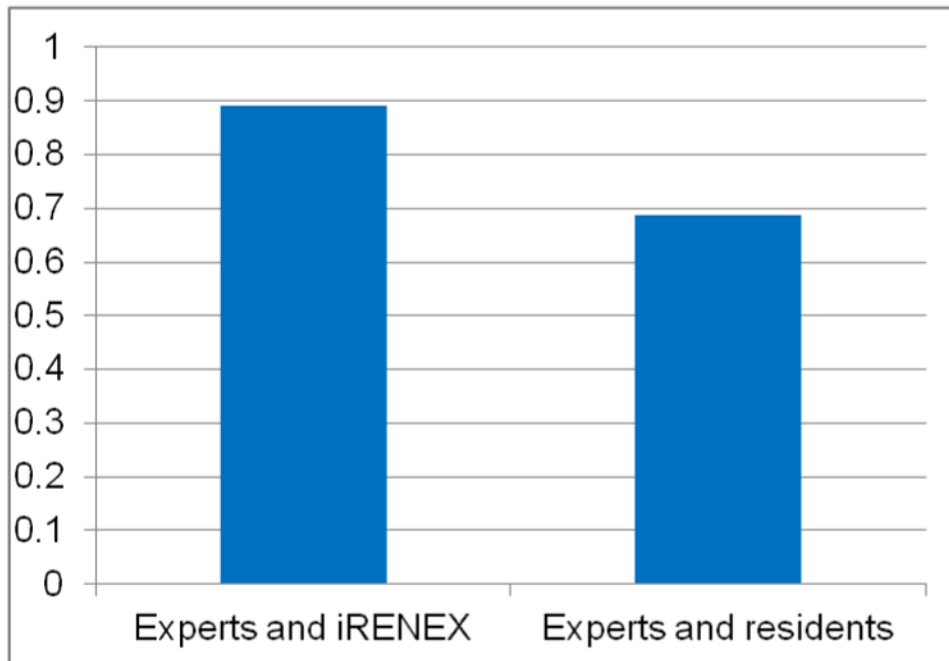
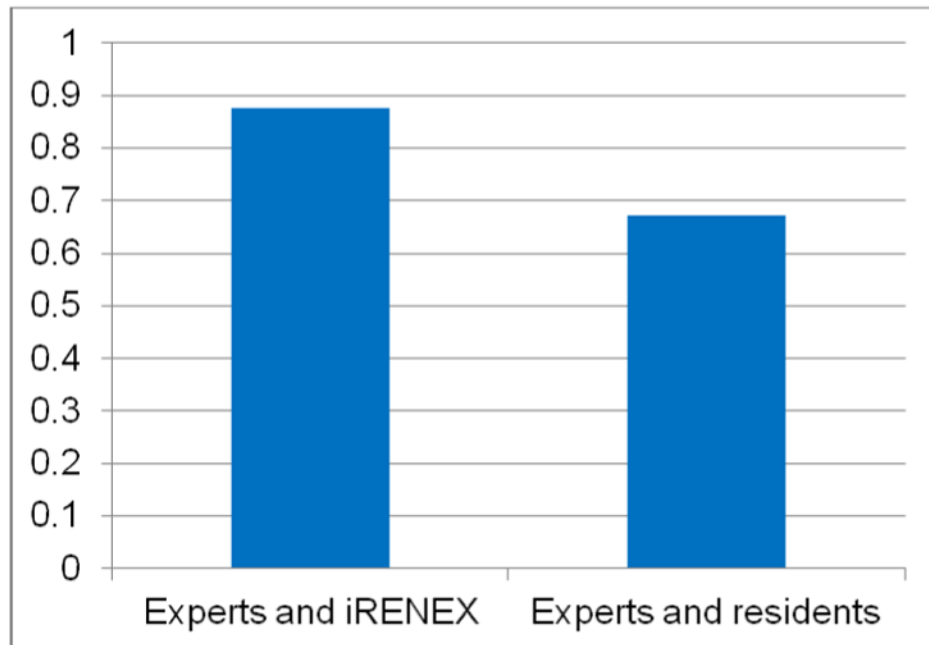


Fig 4.

Correlation between iRENEX certainty factors and the median certainty factors of the three human experts regarding obstruction of the left kidney; comparable results (not shown) were obtained for the right kidney. The red lines mark the equivocal boundaries for certainty factors between -0.2 and $+0.2$. Note the excellent degree of correlation confirming that the conclusions reached by iRENEX have a degree of certainty comparable to those reached by experts (68, reproduced by permission).

5A.



5B.**Fig 5.**

Agreement (concordant correlation coefficient) between iRENEX and experts and residents and experts regarding the presence or absence of obstruction:

5A. CCC* among experts and iRENEX and experts and residents using the ^{99m}Tc -MAG3 scans and renogram data without clinical information

* Concordance correlation coefficient, $p < 0.05$

5B. CCC* among experts and iRENEX and experts and residents using the ^{99m}Tc -MAG3 scans and renogram data with clinical data

* Concordant correlation coefficient, $p < 0.05$

Table 1

Camera-based MAG3 Clearances (ml/min/1.73m²)

	Gender	N	Mean	Std. Dev	Minimum	5 th percentile	95 th 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

* The difference is significant ($p < 0.05$) between males and females. Std. Dev. = standard deviation

Table 2
Post-Void / Maximum Count Ratios Using Regions of Interest over the Entire Kidney*

	N	Mean	SD	Minimum	5 th percentile	95 th 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

* 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. SD = standard deviation

Table 3
Regression equations defining normal limits for the left and right renal area and length*

	5th Percentile	95th Percentile
Left Kidney Area (cm ²)	$32.5 + 9.6*BSA$	$12.6 + 31.7*BSA$
Right Kidney Area (cm ²)	$16.1 + 18.5*BSA$	$32.6 + 22.2*BSA$
Left Kidney Axis (cm)	$8.2 + 1.3*BSA$	$9.1 + 2.3*BSA$
Right Kidney Axis (cm)	$8.8 + 1.0*BSA$	$11.1 + 1.4*BSA$