CLINICAL EVALUATION OF A RENAL MASS DIAGNOSTIC EXPERT SYSTEM

PHEI LANG CHANG,*† YU CHUAN LI,‡ CHI JU WU*
MING HSIUNG HUANG* and PETER JOHN HAUG‡

* Department of Urology, Chang Gung Memorial Hospital, Chang Gung Medical College, Taipei, Taiwan, Republic of China; and ‡ Department of Medical Informatics, University of Utah, Salt Lake City, Utah, U.S.A.

(*Received* 28 *December* 1993; *in revised form* 29 *April* 1994; *received for publication* 13 *May* 1994)

Abstract—In this paper, we describe our clinical evaluation of the diagnostic accuracy of the renal mass diagnostic system (RMDS) and of seven physicians. To investigate the value of intravenous urography (IVU) and/or retrograde urography (RU) in diagnosing renal parenchymal tumors and tumors of the renal pelvis, RMDS and the seven physicians were tested with and without the information regarding IVU/RU at two different times. From this study we believe that RMDS can help residents in making more accurate presurgical renal mass diagnosis, and may eliminate the need for IVU/RU in the diagnosing process for a specific group of patients.

Decision support system Expert system Renal mass Intravenous urography Retrograde urography

INTRODUCTION

The preoperative differential diagnosis of renal masses has been a challenging and expensive process. However, correct preoperative categorization of renal tumor is important because the selection of the operative approach often depends on it. Not only can the positioning of the patient and the extent of incision be different during the procedure, but the aggressiveness of surgical resection can also be different for various types of renal tumors. Radical operation for tumors of the renal parenchyma usually includes the removal of kidney, adrenal gland, perinephric tissue, part of the ureter and regional lymph nodes, while for tumors in the renal pelvis, surgeons often need to remove the whole ureter as well as a cuff of urinary bladder in addition to the kidney and adrenal gland [1–4]. Although the direct observation and frozen sections of tissue during the operation account for part of the evidence in deciding the surgical approach, accurate presurgical diagnosis is important in reducing the uncertainty in this decision process.

The difference in surgical approach makes the determination of renal mass origin (renal parenchyma versus renal pelvis) a crucial decision. Among the procedures that can help in determining the tumor origin, intravenous urography (IVU) and/or retrograde urography (RU) are mainly used to visualize the renal collecting system and are thought to be helpful in identifying renal pelvic tumors. However, it is unclear whether IVU/RU contribute to the diagnoses of renal parenchymal tumors.

To help inexperienced physicians with their diagnosis, we have developed a preoperative renal mass diagnostic system (RMDS) using a Bayesian probabilistic approach [5]. In this paper, we describe a study using 108 renal mass patients in an evaluation of the diagnostic accuracy of RMDS and of seven physicians. We also investigated the value of IVU/RU in identifying renal parenchymal tumors and tumors of the renal pelvis. This

† Author to whom correspondence should be addressed, at: Department of Urology, No. 5, Fu-Shing Street, Kweishan, 333, Taoyuan, Taiwan, Republic of China.

MATERIALS AND METHODS

Structure of the RMDS

RMDS was developed using the ILIAD shell, which is a set of tools best known as the foundation for a large diagnostic system for internal medicine [6, 7]. Diseases are constructed as frames in which the prior probabilities of the diseases and the conditional probabilities for findings are embedded. Several mechanisms including multi-level frames have been implemented in this system to handle conditionally dependent findings [7–10]. When only single-level frames are used in the knowledge base, the system behaves as a multi-membership Bayesian program [7, 8, 11].

The construction of RMDS uses principally the single-level structure. It consists of 18 probabilistic disease frames, classified into five categories, each representing one form of renal mass that can be seen in a urological department (Table 1) [12]. The number of findings per frame ranges from 9 to 23 with an average of 15. Prevalence rates (prior probabilities) of the 18 renal mass diseases were calculated from a large patient database of Chang Gung Memorial Hospital independent of the test cases. The conditional probabilities of findings for each disease were estimated cooperatively by senior urologists.

Patients

There were 111 consecutive patients with renal masses diagnosed and operated on between May 1991 and April 1992 in the section of urology in Chang Gung Memorial Hospital. Three cases were excluded (one case of metastatic tumor of kidney and two cases of renal cell carcinoma) because those patients were found to be uremic and hence IVU could not be performed. In these 108 test cases, 51 were male and 57 were female patients. Patients' age ranged from 6 to 85 years, with an average of 56. The final diagnoses of these 108 cases were all confirmed by pathological examination after operation and the results were used as the gold standard diagnoses in this study. Table 1

Table 1. The distribution of disease for the test cases

1. Renal parenchymal tumors	41 (38%)
Angiomyolipoma	6 (6%)
Hemangiopericytoma	0 (0%)
Juxtaglomerular cell tumor	0 (0%)
Lipoma	0 (0%)
Lymphoblastoma	4 (4%)
Metastatic tumor	1 (1%)
Oncocytoma	2 (2%)
Renal cell carcinoma	24 (22%)
Sarcoma	2 (2%)
Wilms' tumor	2 (2%)
2. Tumors of renal pelvis	39 (36%)
Benign papilloma	3 (3%)
Transitional cell carcinoma	32 (30%)
Squamous cell carcinoma	2 (2%)
Adenocarcinoma	2 (2%)
3. Renal cyst	12 (11%)
Simple cyst	12 (11%)
Cystadenocarcinoma	0 (0%)
4. Renal abscess	11 (10%)
5. Xanthogranulomatous pyelonephritis (XGP)	5 (5%)
Total	108 (100%)

lists the distribution of diagnoses for the test cases. Four kinds of renal tumor were not found in these test cases because of their rarity. There were only two Wilms' tumor patients because children with Wilms' tumor were typically admitted to the section of pediatric surgery. The number of patients with renal cysts also did not correspond to its disease prevalence because few renal cysts required surgical exploration.

In these 108 patients, all received ultrasonography and IVU study. Thirty-three patients had RU study due to poor visualization of IVU. One hundred and two patients had CT scan and six patients with simple cyst of the kidney received percutaneous aspiration instead of CT scan. Thirteen patients had renal angiography combined with CT scan. None of these patients had an MRI examination.

Findings of cases without IVU/RU information were abstracted from the charts and input into RMDS by an independent physician. One urologic young attending physician (Attending), three chief residents of the urologic department (numbered CR-1, CR-2 and CR-3) and three second-year urologic residents (numbered R2-1, R2-2 and R2-3) were presented with the same extracted patient data and asked to state the most likely diagnosis for each case. The diagnoses made by RMDS were considered "correct" if and only if the top diagnosis listed by RMDS matched the gold standard diagnosis and the likelihood of this diagnosis was predicted as greater than 50%.

Analysis

In order to evaluate the importance of IVU/RU in identifying the origin of tumors, positive predictive values (PV+) and negative predictive values (PV−) for the physicians and RMDS were calculated on the basis of correct identification of the tumor category (i.e. a diagnosis is counted as correct if it and the gold standard diagnosis belong to the same category). Three months later, the same set of patient data with IVU/RU information was represented to the physicians and RMDS for diagnosis. The overall diagnostic accuracy, PV+ and PV− were then calculated again. In our further discussion, we label the set of patient data *without* IVU/RU information as *Set 1* and the *complete* patient data set as *Set 2*. In the analysis of PV+ and PV− for the major categories in Table 1, only the first two categories (renal parenchymal tumors and tumors of the renal pelvis) were evaluated because of the relevance of IVU/RU to diseases in these categories.

We used the Wilcoxon signed rank test to measure the statistical significance of the difference of physicians' performance in Set 1 and Set 2. To compare the diagnostic accuracy between RMDS and the physicians, McNemar's test for correlated proportions was applied [13]. The differences of predictive values between RMDS and the physicians were tested using Pratt's approximation of binomial confidence limits [14]. All the comparisons between RMDS and the physicians were adjusted using the modified Bonferroni technique proposed by Keppel [15].

RESULTS

The overall diagnostic accuracy of RMDS and the physicians is shown in Table 2. The diagnostic accuracy for the group of physicians in Set 2 (54.6–77.8%) is significantly better than those in Set 1 (48.1–65.7%, $p < 0.02$). The diagnostic accuracy of RMDS was also compared to each physician in our study in both sets. In Set 1, RMDS was able to diagnose correctly 77 out of 108 patients (71.3%), which is significantly better than every CR (58.3–60.2%) and R2 (48.1–52.8%), while not significantly different from the young attending physician (65.7%). Ninety out of the 108 patients were correctly diagnosed by RMDS in Set 2 (83.3%). This is significantly better than every CR (65.7–70.4%) and R2 (54.6–59.3%), and again, not significantly different from the young attending physician (77.8%).

Tables 3(a) and 3(b) show the performance of the physicians and RMDS in terms of distinguishing a renal parenchymal tumor from a renal pelvic tumor using PV+ and PV−. For a renal parenchymal tumor, IVU/RU information made little difference in

Table 2. The differences of diagnostic accuracy between Set 1 and Set 2 ($N = 108$)

	Set 1	Set 2
RMDS	77 (71.3%)	90 (83.3%)
Attending	71 (65.7%)	84 (77.8%)
CR-1	63 (58.3%)*	71 (65.7%)**
CR-2	65 (60.2%)*	73 (67.6%)*
CR-3	64 (59.3%)*	76 (70.4%)*
R2-1	57 (52.8%)**	64 (59.3%)**
R2-2	52 (48.1%)**	62 (57.4%)**
R2-3	54 (50.0%)**	59 (54.6%)**

Set 1: without IVU/RU information; Set 2: with IVU/RU information. The overall diagnostic accuracy in Set 2 is significantly better than in Set 1 with $p < 0.02$.

* $p < 0.02$ compared to the accuracy of RMDS in the same set.

** $p < 0.001$ compared to the accuracy of RMDS in the same set.

Table 3(a). The positive predictive values (PV+) and negative predictive values (PV−) for renal parenchymal tumor across Set 1 and Set 2

	Renal parenchymal tumor			
	Set 1 (%)		Set 2 (%)	
	PV+	PV−	PV+	PV−
RMDS	88.1	93.9	88.4	95.4
Attending	74.4	86.2	85.0	89.7
CR-1	60.9**	79.0*	75.0	83.8*
CR-2	61.7**	80.3*	76.9	84.1*
CR-3	59.2**	79.7*	78.0	86.6
R2-1	56.0**	77.6**	70.0*	80.9*
R2-2	50.0**	70.3**	61.8**	73.0**
R2-3	50.0**	71.0**	61.1**	73.6**

Table 3(b). The positive predictive values (PV+) and negative predictive values (PV−) for renal pelvic tumor across Set 1 and Set 2

	Renal pelvic tumor			
	Set 1 (%)		Set 2 (%)	
	PV+	PV−	PV+	PV−
RMDS	87.5	85.5	92.5	97.3
Attending	72.7	80.0	82.5	91.2
CR-1	61.8*	75.7	72.5*	85.3*
CR-2	63.6*	76.0	73.2*	86.6*
CR-3	61.3*	74.0	76.9*	87.0*
R2-1	56.7**	71.8*	67.5**	82.4**
R2-2	47.2**	69.4*	56.5**	79.0**
R2-3	47.1**	68.9*	56.8**	78.1**

* $p < 0.02$ compared to the predictive value of RMDS in the same column.

** $p < 0.001$ compared to the predictive value of RMDS in the same column.

PV+ (from 88.1% to 88.4%) and PV− (from 93.9% to 95.4%) when presented to RMDS, whereas in the physician group, a significant increase of PV+ ($p<0.02$) and PV− ($p<0.02$) between Set 1 and Set 2 was observed. In identifying renal pelvic tumors, IVU/RU information made a significant contribution to PV+ ($p<0.02$) and PV− ($p<0.02$) in the physician group and a substantial increase in PV+ (5.0%) and PV− (11.8%) of RMDS.

In comparing RMDS to individual physicians in terms of PV+ and PV−, we found that RMDS was significantly more accurate than every CR and R2 in diagnosing renal parenchymal tumor-Set 1 and renal pelvic tumor-Set 2. In renal parenchymal tumor-Set 2 and renal pelvic tumor-Set 1, RMDS significantly outperformed every R2 and several CRs. None of the comparisons between RMDS and the young attending physician was statistically significant.

DISCUSSION AND CONCLUSIONS

Evaluation of clinical decision support systems is a complex issue [16–18]. In order to understand the effectiveness of the ILIAD Shell knowledge representation, we have begun by assessing the diagnostic accuracy of this prototype system. The clinical area chosen is the differential diagnosis of renal masses. A focus of the study was the differentiation of renal parenchymal and renal pelvic tumors with two subsets of the clinical data. The diagnostic accuracy and PV+/PV− of RMDS were compared to seven physicians with different degrees of training in urology to understand the theoretical usefulness of this system. However, we have not formally evaluated the effort needed for data abstraction and data entry, which have been described as one of the major barriers for physicians' acceptance of decision support systems [19, 20].

As described in the Results section, the overall diagnostic accuracy in Set 2 is significantly better than in Set 1. This is consistent with the general belief that IVU/RU examinations are useful in diagnosing renal masses. The significant difference between the accuracy of RMDS and the six urologic residents suggested that the knowledge representation and inferencing algorithm we chose can achieve potentially useful diagnostic accuracy in this domain.

The PV+/PV− results demonstrate an interesting difference in the way RMDS and the physicians utilized IVU/RU information. In Table 3(a), where the PV+/PV− for renal parenchymal tumors are shown, RMDS did not show much gain in PV+/PV− given the IVU/RU information. Yet the physicians improved their PV+/PV− significantly from Set 1 and Set 2. This discrepancy suggests that the information provided by IVU/RU, while helping physicians in diagnosing renal parenchymal tumor, is not as important to RMDS. We also noticed the trend that RMDS performed significantly better than all residents in Set 1 but not in Set 2.

Table 3(b) shows an evident increase of PV+/PV− for RMDS and the physicians, which suggested the general importance of IVU/RU for diagnoses of renal pelvic tumors. RMDS continued to perform better than all residents. This difference was significant for PV+ in Set 1 and for both PV+ and PV− in Set 2. Although no statistical significance was detected in any comparisons between RMDS and the young attending physician, RMDS exhibited higher PV+/PV− in both sets and tumor categories.

The limitation of the current inferencing algorithm (multi-membership Bayesian) in RMDS may be one reason why, unlike human experts, RMDS showed little improvement in diagnosing renal parenchymal tumors from IVU/RU. When we enter a negative IVU/RU report (which is often the case for a patient with renal parenchymal tumor) into RMDS, the estimated probabilities of renal pelvic tumors will be depressed somewhat while the estimated probabilities of renal parenchymal tumors remain unchanged. However, for a human expert, a negative IVU/RU report not only lowers the likelihood of a tumor in renal pelvis, but also suggests that any tumor present is more likely to be in the renal parenchyma. Researchers in Bayesian diagnostic algorithms describe this capability of redirecting evidence to the other diagnoses when the likelihood of one is

reduced as "*d*-separation". A probabilistic inference scheme called "Bayesian belief network" is believed to handle this problem accurately [16, 21, 22].

Even though we may consider the insensitivity to IVU/RU information a technical drawback of RMDS, one interesting aspect is that a patient diagnosed by RMDS as having renal parenchymal tumor may not have to go through the IVU/RU examination according to the current data. These examinations only increase the RMDS's PV+ by 0.3% and PV− by 1.5%. In Set 1, the PV+/PV− of RMDS were higher than those of the young attending physician in Set 2. A diagnostic inferencing system with this behavior could potentially yield results similar to the unaided physician without incurring the expense of the IVU/RU examination.

The fact that RMDS exhibited highest PV+/PV− across all conditions suggests its potential usefulness as a computerized consulting system if physicians are willing to use it. Better presurgical renal mass diagnoses often result in shorter surgical times, less aggressive operations and hence fewer complications and a reduced length of stay. Nevertheless, a cost–benefit analysis involving patient utilities and cost of the diagnostic and treatment procedures would be necessary to justify the effort needed to construct a diagnostic support system and the medical staff time used to abstract and input patient data. We are currently planning to investigate the impact of these sorts of computerized diagnostic suggestions on the physician's final decision in a prospective study.

We believe that the integration of computerized diagnostic support tools into the clinical settings is the key to maintaining the quality of medical care while managing the costs associated with the diagnosis and treatment of the challenging patients found in the modern hospital environment.

SUMMARY

We have developed a microcomputer-based diagnostic expert system for renal mass differential diagnosis that we call RMDS (renal mass diagnostic system). In this paper, we describe our clinical evaluation of the diagnostic accuracy of RMDS and of seven physicians. We also investigated the value of intravenous urography (IVU) and/or retrograde urography (RU) in identifying renal parenchymal tumors and tumors of the renal pelvis.

In a retrospective study, 108 cases of renal mass were collected for use as experimental subjects. Diagnostic accuracy for these test cases was assessed for RMDS and seven physicians (one attending physician, three chief residents and three second-year residents). To investigate the value of IVU/RU in diagnosing renal parenchymal tumors and tumors of renal pelvis, RMDS and the seven physicians were tested with and without the information regarding IVU/RU at two different times.

RMDS was able to diagnose correctly 90 out of the 108 cases (83.3%), which is not significantly different from the diagnostic accuracy of the attending physician (77.8%). However, this accuracy rate was better than all the other physicians (range from 54.6% to 70.4%). For renal parenchymal tumor, IVU/RU information made little difference in positive predictive value (PV+) and negative predictive value (PV−) when presented to RMDS, whereas in the physician group, a significant increase of PV+ ($p<0.02$) and PV− ($p<0.02$) was observed. In identifying renal pelvic tumor, IVU/RU information made a significant contribution to PV+ ($p<0.02$) and PV− ($p<0.02$) in the physician group and a substantial increase in PV+ (5.0%) and PV− (11.8%) of RMDS.

We believe that RMDS can be used as a computerized consultant that helps physicians in making more accurate presurgical renal mass diagnoses. For a specific group of patients, the use of this sort of computerized assistance may eliminate the need for IVU/RU in the diagnosing process.

Acknowledgements—We are indebted to Dr Homer R. Warner and other members of the Department of Medical Informatics at University of Utah, for helpful suggestions concerning the manuscript and the HELP shell systems that made this project possible. We also thank Dr John Kircher in the Department of Psychology at University of Utah for statistical consultation.

REFERENCES

1. C. J. Robson, Radical nephrectomy for renal cell carcinoma, *J. Urol.* **89**, 37–42 (1963).
2. K. B. Cumming, Nephroureterectomy: rationale in the management of transitional cell carcinoma of the upper urinary tract, *Urol. Clin. North Am.* **7**, 569–578 (1980).
3. C. B. Williams and J. P. Mitchell, Carcinoma of the ureter: a review of 54 cases, *Br. J. Urol.* **45**, 377–387 (1973).
4. E. A. Tanagho and J. W. McAninch, *Smith's General Urology*, 13th Edn, p. 359. Prentice-Hall, San Mateo, CA (1992).
5. P. L. Chang, Y. C. Li, C. J. Wu and M. H. Huang, Using Iliad system shell to create an expert system for differential diagnosis of renal masses, *J. med. Syst.* **17**, 289–297 (1993).
6. R. Cundick, C. W. Turner, M. J. Lincoln, J. P. Buchanan, C. Anderson, H. R. Warner Jr and O. Bouhaddou, Iliad as a patient case simulator to teach medical problem solving. *Proceedings of the 13th Symposium on Computer Applications in Medical Care (SCAMC)*, p. 902. IEEE Computer Society Press, Washington DC (1989).
7. C. W. Turner, M. J. Lincoln, P. J. Haug, H. R. Warner, J. W. Williamson and N. Whitman, Clustered disease findings: aspects of expert systems, *International Symposium of Medical Informatics and Education*, p. 259. University of Victoria, B.C., Canada (1989).
8. H. Yu, P. J. Haug, M. J. Lincoln, C. Turner and H. R. Warner, Clustered knowledge representation: increasing the reliability of computerized expert systems, *Proceedings of the 12th Symposium on Computer Applications in Medical Care (SCAMC)*, p. 126. IEEE Computer Society Press, Washington DC (1988).
9. D. K. Sorenson, R. M. Cundick, C. Fan and H. R. Warner, Passing partial information among Bayesian and Boolean frames, *Proceedings of the 13th Symposium on Computer Applications in Medical Care (SCAMC)*, p. 50. IEEE Computer Society Press, Washington DC (1989).
10. M. J. Lincoln, P. J. Haug, Y. Hong, C. Turner and H. R. Warner, Expert biases prevent accurate estimation of population statistics for clustered disease frames, *International Symposium of Medical Informatics and Education*, p. 237. University of Victoria, B.C., Canada (1989).
11. H. C. Sox, M. A. Blatt, M. C. Higgins and K. I. Marton, *Medical Decision Making*. Butterworth Heinemann, Boston (1988).
12. W. H. Lakey, *Tumors of the Kidney, Urology*, Vol. 2. Harper & Row, New York (1975).
13. Q. McNemar, *Psychological Statistics*, 4th Edn, p. 54. Wiley, New York (1969).
14. C. Blyth, Approximate binomial confidence limits, *J. Am. stat. Assoc.* **81**, 843–855 (1986).
15. G. Keppel, *Design and Analysis: A Researcher's Handbook*, 3rd Edn, p. 169. Englewood Cliffs, Prentice-Hall, NJ (1991).
16. Y. Li and P. J. Haug, Evaluating the quality of a probabilistic diagnostic system using different inferencing strategies, *Proceedings of the 17th Symposium on Computer Applications in Medical Care (SCAMC)*. Washington DC (1993) (in press).
17. J. Hilden and J. D. F. Habbema, Evaluation of clinical decision aids—more to think about, *Med. Inform.* **15**, 275 (1990).
18. P. L. Miller, Issues in the evaluation of artificial intelligence system in medicine, *Proceedings of the 9th Annual Symposium on Computer Applications in Medical Care (SCAMC)*, p. 281. IEEE Computer Society Press, New York (1985).
19. E. H. Shortliffe, Computer programs to support clinical decision making, *J. Am. med. Assoc.* **258**, 61–66 (1987).
20. W. M. Tierney, M. E. Miller, J. M. Overhage and C. J. McDonald, Physician inpatient order writing on microcomputer workstations, *J. Am. med. Assoc.* **269**, 379–383 (1993).
21. J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufman, San Mateo, CA (1988).
22. E. Neopolitan, *Probabilistic Reasoning in Expert Systems*. Wiley, New York (1990).

About the Author—Phei Lang Chang, M.D., is the Chairman of the Department of Urology, Chang Gung Memorial Hospital, Chang Gung Medical College, Taipei, Taiwan, Republic of China.

About the Author—Yu Chauan Li, M.D., is a Ph.D. candidate at the Department of Medical Informatics, School of Medicine, University of Utah, Salt Lake City, Utah, U.S.A.

About the Author—Chi Ju Wu, M.D., is an Attending Doctor at the Department of Urology, Chang Gung Memorial Hospital, Chang Gung Medical College, Taipei, Taiwan, Republic of China.

About the Author—Ming Hsiung Huang, M.D., is an Attending Doctor, Department of Urology, Chang Gung Memorial Hospital, Chang Gung Medical College, Taipei, Taiwan, Republic of China.

About the Author—Peter John Haug is an Associate Professor, Department of Medical Informatics, School of Medicine, University of Utah, Salt Lake City, Utah, U.S.A.