大韓核醫學會誌：第20卷 第2號

Statistical Analysis of the IAEA-WHO Liver Phantom Images for the Asian Countries

Seong Woon Hong, M.D.

Department of Nuclear Medicine, Cancer Center Hospital

Tae Woong Kang, M.D.

Department of Internal Medicine

＝국문초록＝

아시아국가에서 IAEA-WHO 간모형 영상처리에 관한 통계학적 고찰

원자력병원 핵의학과

홍 성 운

내 과

강 태 웅

핵의학기기중 scintillation gamma camera의 임상적 유용도는 이미 잘 알려져 왔으며 80년대에 들어와서는 computer의 도입으로 그이용도가 더욱 확대되어 보유댓수가 급격한 증가를 이루게 되었다. 따라서 이에 대한 정도관리가 필요하게 되었다. 정도관리의 목적은 장비기능의 정상여부를 조기에 발견하여 항상 균등한 질의 영상을 재현하여 보다 정확한 진단을 하는데 있다. 따라서 r-camera의 사용자는 기계의 기능과 성능을 항상 정확하게 파악하여 빠르고 적절한 대책을 세워 양질의 영상을 얻도록 해야한다. 또한 스캔의 결과도 판독자 환자대상군, 검사방법에 따라 스캔의 예민도 및 특이도가 달라지며 정확도 또한 공간점유병소의 위치, 원인, 병소수의 크기에 따라 달라진다고 하였다. 저자는 이점을 감안하여 대상군, 검사방법 병소위치에 의한 변화를 배제하고 판독자의 검출정확도를 알아보고자 IAEA-RCA 협조를 얻어 IAEA-WHO 제공 간모형(SALP: simulatied anatomic liver phantom)을 사용하여 국내 16개 병원 핵의학과 staff 20명에게 의뢰하여 얻은 결과와 아시아 8개국의 276명의 결과를 함께 분석하여 다음과 같은 결론을 얻었다.

1) 각개인의 간모형영상의 판독 정확도는 60%～100%사이었고 대부분 90% 내외였다.
2) 아시아 태평양지역의 정확도는 유럽 및 라틴아메리카의 결과와 비슷하였다.
3) 각 나라별의 정확도 결과는 91.1%에서 76.4%를 나타내었다.
4) 스캔너를 사용한 영상의 판독결과와 감마 카메라의 영상판독결과는 차이가 없었다.
5) 정도관리빈도와 정도관리검사방법은 영상판독 정확도의 결과와는 무관 하였다.

*이 연구는 IAEA-RCA 사업의 연구보조비로 이루어 진 것임.

Introduction

The application of nuclear meidcine to clinics become generalized and got the important position in field of diagnosis.

The develpoment of nuclear medicine instrument has grown rapidly in 1980's especially the gamma camera are discussed as the basis for quality control.

Quality control of gamma camera is necessary to insure maximum diagnostic information at the minimum possible risk of patients.

The purpose of the survey is to inform paticipation laboratories as to the quality of the results of their imaging procedures.

Quality control study method employed simulated anatomic liver phantom (SALP) in each institute were analysed. The SALP is identical to that provided from IAEA to paticipation laboratories.

The image of phantom by the procedure which normally use for liver imaging with ^{99m}Tc labelled radiopharmaceuticals, and to evaluate the resultant image for any areas which appear abnormal.

Findings are to be recorded on the accompanying evaluation from, proceded by details of the imaging procedure.

Materials and Method

1. Imaging of SALP

The SALP is a transimssion phantom and contains no radiocativity. In order to obtain an image, it must be covered by a uniform extended radioactive source emitting photons below 150 keV in energy. A fillable flood phantom containing 75-100 MBq (2-3 mCi) of ^{99m}Tc is preferable, however an extended plastic discsource containing a similar amount of ^{57}Co may be used.

2. Procedure with scintillation camera

1) Choice of collimator

Mount the collimator normally used for static liver imaging on the detector head.

2) Positioning of the SALP

Position the SALP as in Figure 1 with the anterior surface adjacent to the collimator and the posterior surface covered with the radioactive source. The SALP should be oriented so that its cephalad edge corresponds to the upper margin of the image and vice versa.

The SALP will not cover the entire field of view. No radiation reaches the detector except that which has traversed the body of the SALP. This is done by placing lead sheets or aprons just under all four edges of the phantom, covering the exposed area of the collimator.

3) Choice of imaging parameters

a) Photon energy and window width

Adjust the spectrometer for 122 keV if using a ^{57}Co source of 140 keV if using ^{99m}Tc.

Adjust the window width to that normally used for liver imaging.

b) Other parameters

Obtain an image on the hard copy device normally used for clinical imaging, making certain that the image size, film type and number of acquired counts are the same as used for patients. If computer processing is used, the same processing techniques as applied to clinical images should be em-

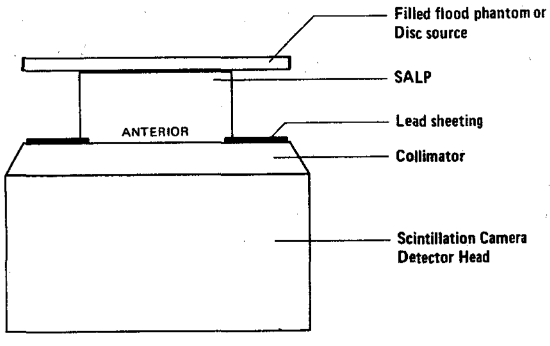

Fig. 1. Detector Head-Phantom-Source Configuration for imaging SALP.

ployed. If alternative acquisition or processing modes are to be investigated, a separate evaluation form must be completed for each.

3. Procedure with rectilinear scanner

1) Choice of collimator

Mount the collimator normally used for liver imaging on the detector probe.

2) Positioning of SALP

Position the SALP as in Figure lb such that the anterior surface is closest to the collimator and the posterior surface is covered with the radioactive source. Invert the phantom-source configuration if the scanner head is directed upward. Consider the mid-plane of the lesions to be 5 cm below the anterior surface of the phantom (Fig. 2).

3) Choice of imaging parameters

a) Photon energy and window width

Adjust the spectrometer for 122 keV if using a ^{57}Co source of 140 keV if using ^{99m}Tc.

Adjust the window width to that normally used for liver imaging.

4. Evaluation of SALP Image

The evaluation grid and the coordinate conventions are shown and, most importantly, the exact locations of the lesions contained in the phantom. The figures printed in bold in the corrdinate grid are their rations. The rating scale has been

1-for definitely no lesion,

2-for probable no lesion,

3-for probably a lesion and

4-for definitely a lesion.

The optimum result achievable would therefore be a 1 in all quadrants without a lesion and a 4 in all quadrants containing a lesion. If the evaluation contains several false positive or negative findings it is worthwhile to check the pattern observed against the uniformity of the imaging device.

The method of statistical information employed is based on the generation of an ROC-curve for each individual participant. The ROC-curve is constructed according to the method of graded rating. A description of this method including practical samples can be found in Hanley JA, McMeil Barbara J: The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve. Radiology 143, 29-36, 1982. The ROC Curve obtained from the rating results of all participants is shown below. The area under the ROC-curve is a numerical value that reflects the quality of the performance. In the

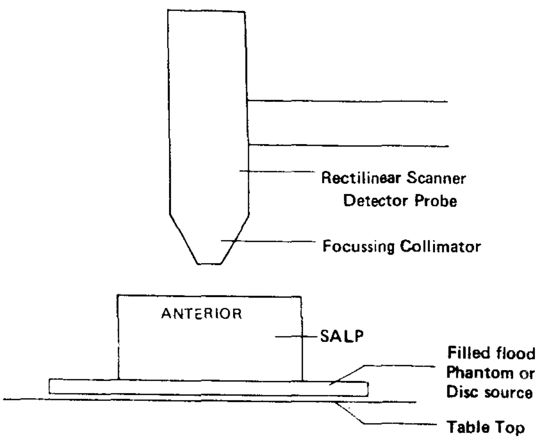

Fig. 2. Recilinear Scanner Detector Probe-Phantom-Source Configuration for imaging SALP.

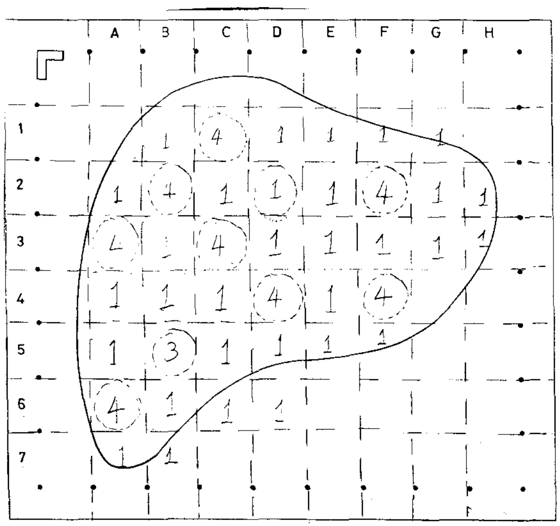

Fig. 3. Grid system to be used to define presence or absence of abnormalities in SALP image.

case of the group ROC the area is 87.6% of the total area of the square. The performance of an imaging thest is the better the higher the area under the ROC-curve. The maximum area that can be reached is 100%, however, anyone obtaining this result has had almost unbelievable luck.

The ROC-areas have been used to rank the individual results. Since more than 200 participants have submitted valid results we have been able to perform simple statistical evaluations. The most straightforward evaluation has been to generate a histogram of the ROC-areas (Fig. 3).

Result

A total of 95 institutions from 9 Asian countries had returned the results of their simulated liver transmission phantom evaluation. There were on an average 3 evaluation sheets per institute. A total of 122 imaging devices were included in the survey, 110 of them were gamma cameras and 12 were scanners.

The following table summarizes the data of the participating countries, the number of institutes per country, the number of evaluation sheets and the number of observers entered in the analysis.

Country	No. of Institues	No. of evaluation sheets	No. of observers
Bangladesh	5	21	10
India	8	51	12
Japan	40	115	75
Korea	16	21	20
Pakistan	7	16	16
Philippines	8	10	13
Singapore	1	1	4
Thailand	8	18	24
Vietnam	2	16	3
Total	95	269	117

1. Analysis of evaluation sheets

1) Methods used

The evaluation of the IAEA-WHO liver phantom image is performed with the aim to distinguish between normal and abnormal areas in the image and to classify the quality of an individual evaluqkation. In prevoius surveys the quality of the imaging procedure has successfully been expressed by counting the number of true positive (TP) and of false positive (FP) decisions. This method has been used by most of the surveys of the college of American Pathologists (Hermann 1977, Herrera 1981).

However, since the reader of an image can-depending on a particular diagnostic situation and on the reader's preferece-apply verying decision thresholds, the number of TPs and TNs found in an image depend on theis decision threshould:

A person who wants to detect as many lesions as possible in an image, will apply a lenient decision threshould and "over-read" the image. This will result in a high number of true positive findings but will simultaneously and unavoidably increase also the number of false positive findings. On the other hand, if the reader wants to keep the false positive rate as low as possible and to cetect only true positive lesions, he will "under-read" the image. This will bring down the number of FPs but will again unavoidably also reduce the number of true positive findings since the possibility of overlooking a ture lesion will be increased. Therefore the number of TPs and FPs depends on the particular decision threshould and judgement of the quality of performance based on these numbers fuffers from the fact that the individual decision threshould is not known. More subtle investigations of e.g. the influence of instrument parameters on the quality of an evaluation are therefore not possible.

A more adequate procedure for image analysis has been investigated and introduced by the college of American Pathologists for their most recent quality control surveys (Hermann 1982). It is based on the use of the Receiver Operating Characteristc (ROC) curve, which is a function plotting the false positive probability against the true positive probability of lesion detection in the phantom. Different points of the curve are obtained by different decision thresholds when reading an image repeatedly. In this survey the variation of the decision threshould is obtained implicitly by using the graded rating method. A more detailed description of this method has been given by e.g. Metz 1978. The graded rating method has been incorporated into the evaluation procedure of the IAEA-WHO liver phantom images by asking the participants to evaluate each square of the test image (in total 41 squares in the image) according to a rating scale with 4 alternative choices (lesion definitely absent, lesion probably absent, lesion probably present, lesion definitely present). Due to the graded rating it is then possible to construct an ROC-curve for each individual evaluation.

An accuracy index can be readily derived from the ROC-curve. It has been demonstrated (Hermann 1982) that the area under the ROC-curve is best surted to indicate the accuracy of such an image evaluation. The ROC-curve as displayed in fig. 4. is typical for the ROC-curves obtained in the survey. The points obtained are always above the diagonal from the origin of the plot to the upper right corner of the plot. The diagonal is termed the guess-line, because, taken as an ROC-curve, the false positive rate always equals the true positive rate. A diagnostic system with such an ROC-curve would describe a test procedure without and diagnostic information. Hence the area of nontrivial ROC-curve is always greater than 0.5 or, al ternatively, 50 percent of the area of the square.

The accuracy of an evaluation is the better the greater the area under the ROC-curve. The possible areas range from 50 percent (no diagnostic information at all in the image) up to 100 percent, the latter

Table 1. Statistics of Individual ROC-areas

Variable	N	Mean	Standard deviation	Minimum value	Maximum value
Area	269	88.96	9.24	49.70	100.00
SDEV	296	6.57	3.23	0.00	11.20

*Area indicates the area under the ROC-curve,
SDEV the standard deviation of the ROC-area obtained according to Hanley-McNeil.

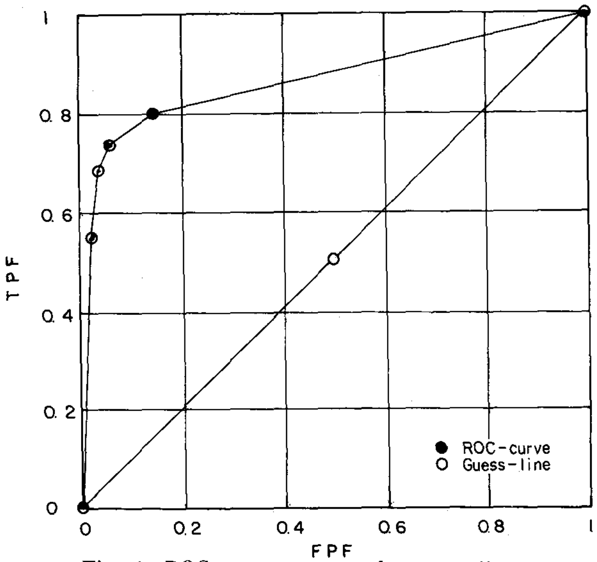

Fig. 4. ROC-curve as opposed to guess-line.

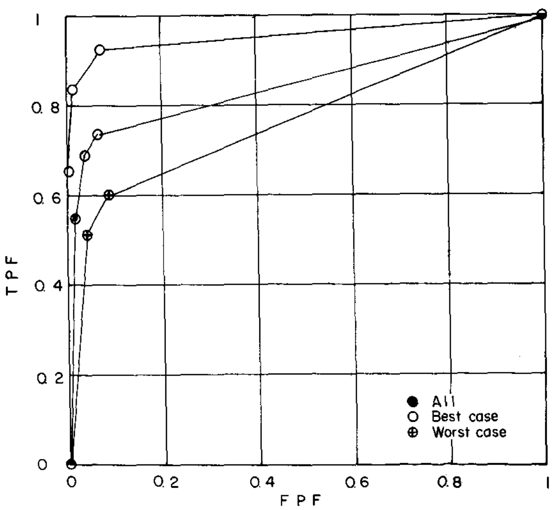

Fig. 5. ROC-curves for SALP evaluation.

representing the perfect diagnostic test with perfect discrimination between abnormal and normal.

The actual calculations of the ROC-areas in this survey have been performed by an algorism giver by Hanley 1982, in which the trapezoidal rule is used to calculat the area from the experimental points.

Group results were obtained by pooling all rew data of the group to construct the group ROC curve.

2) Results

195 individual results have been processed. For each evaluation a separate ROC-curve was generated and the area under was calculated according

Table 2. Histogram of Individual ROC-areas

All ROC-areas are included in the histogram. Areas below 60 percent are put into the class cented at 60.

Midpoint Area	Frequency Bar Chart	FREQ	CUM. FREQ	Percent	CUM. Percent
60.0	* *	4	4	1.49	1.49
62.5	*	1	5	0.37	1.86
65.0	* *	3	8	1.12	2.97
67.5	* *	3	11	1.12	4.09
70.0	*	2	13	0.74	4.83
72.5	* *	4	17	1.49	6.32
75.0	* *	4	21	1.49	7.81
77.5	* * * * * *	12	33	4.46	12.27
80.0	* * * * * * * * *	18	51	6.69	18.96
82.5	* * * * * * * *	16	67	5.95	24.91
85.0	* * * * * * * * * * * * * *	27	94	10.04	34.94
87.5	* * * * * * * * *	18	112	6.69	41.64
90.0	* *	47	159	17.47	59.11
92.5	* * * * * * * * * *	19	178	7.06	66.17
95.0	* * * * * * * * * * * * * * * * * * *	37	215	13.75	79.93
97.5	* * * * * * * * * * *	21	236	7.81	87.73
100.0	* * * * * * * * * * * * * * * * *	33	269	12.27	100.00
	10 20 30 40 Frequency				

Table 3. ROC-areas Grouped into Paticipating Countries, Sorted by Decreasing Area

Country	ROC-area	SD
#1	91.1	1.6
#2	90.8	1.8
#3	90.8	0.6
#4	89.0	8.0
#5	87.5	2.5
#6	86.3	1.2
#7	86.2	1.8
#8	84.4	2.2
#9	76.4	2.1

Table 4. Areas under ROC-curves for Selected Groups

Country	No. of evaluation	ROC-area	SD
ALLASIA	269	87.9	0.5
CAMASIA	200	88.0	0.5
CAMGT80	112	87.7	0.7
CAMLE80	72	88.4	0.9
SCANASIA	53	86.3	1.1

*ALLASIA; all asian results
CAMASIA; results obtained with gamma cameras
CAMGT80; result from gamma camera with year of manufacture after 1980
CAMLE80; results from gamma camera with year of menufacture before and during 1980
SCANASIA; results for scanners.

to the Hanley-McNeil method.

As an example for the graphic representation of ROC-curve three ROC-curve are displayed in fig. 5.

The curve obtained by pooling all results is compared with the best and the worst country result in this survey. The significant differences in performance are clearly demostrated.

The salient conclusion of the analysis were.

i) The individual institution results varied from 60% to 100% but the largest number of results were around 90% (Table 2).

ii) The grouped country-wise results in this region varied from 91.1% to 76.4% (Table 3).

iii) There was no significant difference between the results obtained with scanners and those obtained with gamma cameras (Table 4).

iv) There was no distint correlation eith the type and the frequency of quality control procedures and the results obtained on image analysis.

Conclusions

1) The individual institution results varied from 60% to 100% but the largest number of results were around 90%.

2) The Asia and Pacific results were comparable to the results obtained in a similar survey in Latin America and in Europe.

3) The grouped country-wise results in theis region varied from 91.1% to 76.4%.

4) There was no significant difference between the results obtained with scanners and those obtained with gamma cameras.

5) There was not distinct correlation with the type and the frequency of quality control procedures and the results obtained on image analysis.

REFERENCES

1) Barbara J, McNeil: *Guidlines for evaluating New tests In: Textbook of Nuclear Medicine Lea & Febiger Philadelphia 473, 1979*
2) Drum DE, Beard JO: *Hepaic scintigraphic changes associated alcoholism. J Nucl Med 19:154, 1978*
3) Hanley JA, McNeil BJ: *The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143:29, 1982*
4) Hermann GA, Herrere NE, Hauser W, Oxley DK: *Rationale techniques and results of a quality control program of imaging procedures. In; medical Radionuclide Imaging. IAEA Vienna 1:55, 1977*
5) Hermann GA, Herrera NE, Sugiura HT: *Comparison of inter laboratory data in terms of receiver operating character-istic (ROC) indices. J Nucl Med 23:525, 1982*
6) Herrera NE, Hermann GA, Hauser W, Paras P: *College of American pathologists program series X survey program. In: Medical Radionuclide imaging 1980*
7) Herrera NE, Hermann GA, Hauser W, Paras P: *College of American pathologist program series X survey program. IAEA Vienna 2:177, 1981*
8) Katz MA: *A probability graph describing the predictive value of a highly sensitive diagnostic test. N Engl J M 291:115, 1974*
9) MeNeil BJ, Keeler E, Adelstein SJ: *Primer on certain elements of medical decision making. N Engl J Med 293:211, 1975*
10) Metz CE: *Basic principles of ROC analysis. Semin Nucl Med 8:283, 1978*
11) Metz CE, Kronman HB: *Statistical significance tests for binormal ROC curves. J Meth Psych 22:218, 1980*
12) Rossmann K, Goodenough DJ, Lusted LB: *Radiographic applications receiver operating characteristic (ROC) curves. Radiology 110:89, 1974*
13) Swets JA: *ROC analysis applied to the evaluation of medical imaging techniques. Invest. radio. 14:109, 1974.*
14) Swets JA, Pickett RM, Whitehead SF, et al: *Assessment of diagnostic technologies. Science 205:753, 1979*