Radiation Protection Dosimetry
Vol. 80, Nos 1–3, pp. 73–76 (1998)
Nuclear Technology Publishing

EUROPEAN SURVEY OF IMAGE QUALITY ASSESSMENT METHODS USED IN MAMMOGRAPHY

F. R. Verdun†, R. Moecklì†, J.-F. Valley†, F. Bochud†, C. Hessler‡ and P. Schnyder‡
†Institute of Applied Radiophysics, University Centre, CH-1015 Lausanne, Switzerland
‡University Hospital Centre (CHUV), Department of Diagnostic and Interventional Radiology
CH-1011 Lausanne, Switzerland

Abstract — The definition of reference dose levels has to be linked with the definition of image quality. Unfortunately, there is still no general agreement on the definition of image quality in mammography, and most of the protocols used are based on the detectability of objects having various shapes and contrasts. To facilitate the task of assessing image quality, scoring methods are often used to produce a single number representative of the imaging chain performance. The goal of this study is to present a comparison between different ways of assessing image quality commonly used in Europe. A set of five mammograms, having different image quality levels, has been obtained with several test objects and compared. The results show large sensitivity variations among the different methods. Concerted work between radiologists and physicists is still required to define the radiological tasks and develop objective ways to measure image quality in mammography.

INTRODUCTION

The efficient use of ionising radiation in radiodiagnosis involves three factors: radiographic technique, patient dose and image quality. A good radiograph should produce an image containing all the information essential for diagnosis and should result in the minimum possible dose to the patient[1,2]. To achieve such requirements, careful training and quality assurance programmes need to be organised. In the case of mammography a stricter optimisation scheme needs to be applied since it is one of the most technically demanding procedures, and since mass screening campaigns are now becoming common practice. The quality of radiographic technique relies mainly on radiographers' training. At the present time each country has its own education scheme, but from the growing number of seminars or refresher courses dedicated to this matter, one can expect a continuous improvement of the present situation in the next few years. Concerning patient dose, despite the fact that it can be assessed by means of several quantities, it is always possible to compare different situations. Definition and calibration procedures exist which make conversion tasks possible. The only major problem which remains unsolved is the conversion of the breast dose to the associated risk[3]. Nevertheless as shown by Law[4], more and more epidemiological data can be used to improve the present situation.

Image quality is a well defined subject in imaging sciences. The basic quantities include the Modulation Transfer Function (MTF), the Wiener spectrum, the dynamic range and the contrast. These parameters can be measured separately, but their interplay on the visualisation task of objects is not yet fully understood. This last point is crucial since the efficiency of radiology depends directly upon visualisation tasks, and it is well known that the human brain allows, to a certain extent, the compensation of one image quality parameter by another one. To complicate the situation, visualisation is also observer dependent.

The goal of this study is to present a comparison between different ways of scoring image quality. Thus, images with various image quality levels, were prepared with the common test objects used in Europe, and sent to centres familiar with the test object under study for image quality assessment. The sensitivity of the different scoring methodologies is reported.

MATERIAL AND METHOD

All the test objects involved in the study were imaged under five different imaging conditions in order to obtain a wide range of image qualities varying, as much as possible, each parameter individually (e.g. contrast, resolution and noise).

All the films were obtained using a Siemens Mammomat II unit, which complied with the CEC recommendations[5]. The two different screen–film systems used in the study were the Fuji Fine/Fuji UMMA, and the Kodak MinR/Kodak MinRH system. The standard exposures were made at 28 kV, using an antiscatter grid, a tungsten anode, a molybdenum filter and the Fuji system (Ref. film). Film net optical densities were in the range of 1.1 to 1.4. Film processing was performed in a Typon chemistry (Typon, Switzerland) using a Kodak M35 daylight processor (140 s, 34°C) or an automatic Kodak M8 processor (90 s, 35.5°C).

In order to get a wide range of image qualities, the following imaging conditions were used to expose each test object. Film C– (film with a lower contrast than Ref. film) was obtained with the reference system but at 35 kV, whereas film C+ (films with a higher contrast than Ref. film) was obtained with the reference system but at 25 kV. Film R– (film with a lower resolution and a lower noise level than film Ref.) was obtained with the reference system but with a transparent foil (0.1 mm

thick) sandwiched between the screen and the film, and film N+ (film with a higher noise level than film Ref.) was obtained with the Kodak system and a processing time of 90 s in an inappropriate chemistry.

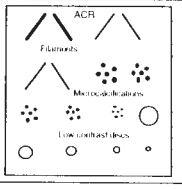
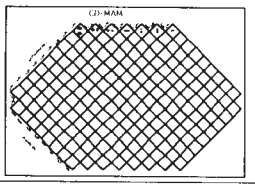
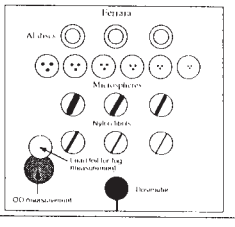
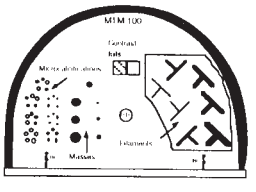
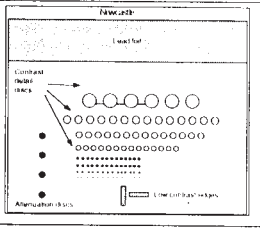
The different test objects involved in the study and the scoring methods used are briefly described in Figure 1. Two methodologies (IQI and NIJ evaluations) used a fully objective assessment of image quality (i.e. no observer involvement).

A set of five films, obtained in the imaging conditions just described, was prepared with each test object. The participants then received the films of their test objects, and were asked to score the different films. The image quality was assessed by at least four different observers in each centre.

To characterise objectively the image quality of the films, three parameters (i.e. contrast, resolution, and noise) were measured using the items available in the IQI test object[6,7]. The films were digitised with an ACS 100 Agfa scanner having a nominal resolution of 2400 dpi (dots per inch) and a dynamic range of 12 bits. The results obtained for the reference film are summarised in Table 1 where C is the contrast between a 0.2 mm thick aluminium foil imbedded in 44.8 mm of PMMA and 45 mm of PMMA (measured in optical den-

Table 1. Objective characterisation of the films.

Film ref.	C (o.d.)	G^2 (μm)	R (mm^{-1})
Ref.	0.30 ± 0.01	2.9 ± 0.2	4.2 ± 0.1
C+	+10%	=	=
C–	–30%	=	=
R–	=	–40%	–48%
N+	=	+158%	=

Test object - scoring method	Designation	Test - object
RMI 156 The final score is the sum of the following numbers assigned respectively for nylon filaments, microcalcifications and masses : 1; 2; 5; 10; 19; 29 - 1; 2; 8; 15; 25 - 1; 2; 3; 10; 20	ACR	
CD-MAM The pseudo-objective score (IQF) is calculated as follows : IQF = $\sum$Cj Dj, where C is the thickness, D the diameter of the disks and j the last disc visible for each thickness.	CD-M	
Ferrare* One is given for each detail seen, 0.5 is given for uncertain detail , and 0 is given if the detail is not seen. The score is the sum of all the numbers produced	FER	
Medi-Test The final score is the sum of the microcalcifications, masses and filaments scores.	MTM	
Newcastle* The score of the mammogram is obtained by adding the number of discs seen and the number of bars seen on the line pairs object (*) not commercially available	NEWC	

Fully visible group	Partly visible group	Score	Score
Mi, Ma, Fi	Mi, Ma, Fi		
1		1	
2		2	3
3	3	4	
4	4	8	6
5	5	16	12
6	6	32	24
7	7		48

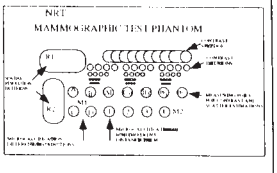
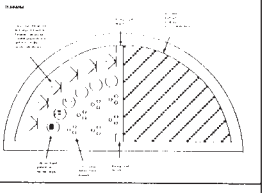
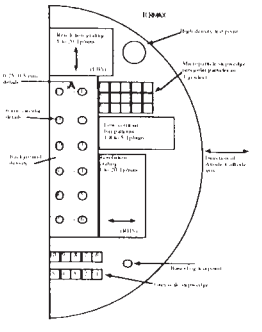
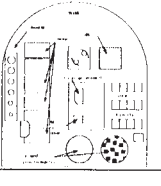
Test object - scoring method	Designation	Test - object
Nordisk Røntgen Teknik The scoring begins with evaluation and a ranking of contrast, spatial resolution, scatter-to primary ratio and base +fog parameters. A weight for each rank is then given: 0.3 for contrast and resolution and 0.2 for S/P ratio and base + fog. The sum is calculated	NRT	
Leeds TORMAM Visibility is graded from 0 to 5 Since there are 6 detail contrasts up to 6 distinct levels of visibility may be perceived. To accomodate this range, it is necessary to use half-scores, plus zero for undetected details	TMAM	
Leeds TORMAX The scores used in this study is the one used by a particular centre. The scores for the low contrast grating, the 6mm details, the 0.5mm detail and the 0.25mm details were simply added. 1 point was awarded in b+f was less than 0.16, 0 if it was between 0.17 and 0.2 and - 1 point if it was more than 0.2. 0 point was then awarded if the contrast index (CI) (which was defined as the difference in optical density between step 9 and step 6) fell between 0.40 and 0.44, 1 point if it fell between 0.45 and 0.49, 2 points if it fell between 0.50 and 0.54 and 3 points if it fell between 0.55 and 0.60. The ranges were chosen so that all films in the trial could be scored	TMAX	 See CD-MAM
CD-MAM The film is digitised and a programme evaluates the percentage of all the disks correctly found.	NIJ	See CD-MAM
Modified Kodak* An image quality index is calculated by means of the pseudo ideal observer. The IQI corresponds to the size of the smallest microcalcification one should be able to detect on a film with a probability of 2% false positive.	IQI	

Figure 1. Test objects and scoring methods used in the survey.

sity unit), G^2 is the Selwyn coefficient, proportional to the zero spatial frequency value of the Wiener noise spectrum, and R is the integral under the MTF curve from 0 to 10 cycles.mm^{-1} (equivalent passband).

RESULTS AND DISCUSSION

To reduce inter-observer variation, the average scores, within centres, of the five films are summarised in Table 2. The first two columns give the results produced by the fully objective methods employed in this study. Several centres did not calculate a score, but gave the results of several parameters available from the test objects. In these cases (i.e. for NRT and TMAX test objects) the scores were calculated by means of the methodology used in a particular centre (see Figure 1). The last line of the table gives an indication of the range of the acceptable scores (when available). The relative differences between the scores of the reference film and the ones produced by the other films have been calculated for each test object. The results have been then categorised in four classes: A sign '+' was given when the relative difference in scores was below 10%; a sign '++' was given when the relative difference in scores was within 11 to 20%; a sign '+++' was given when the relative difference in scores was within 21 to 50%, and a sign '++++' was given when the difference in scores was higher than 51%. These results are reported in Table 3.

A wide range of sensitivities (i.e. relative difference in scores) appears when using the test objects and scoring methods described in Figure 1. Among the two fully objective methods, the IQI is the most sensitive method. However, its sensitivity is relatively low in comparison with other common scoring methods involving observers. The most sensitive methodologies are the ones which used the MTM and ACR methods (i.e. test objects and scoring procedure). However, it should be pointed out that the ACR procedure would have accepted film R–, since its score is above the limit value of 21, in spite of the fact of its lack of resolution. The most surprising results have been obtained with the TMAX test object, which produced large variations in results when considering individual parameters (see Table 4), but which had no sensitivity when applying the scoring methodology. This result shows clearly that the scoring method used here for the TMAX test object was inappropriate.

Table 4. Results produced by the TMAX test object.

Film ref.	Contrast index	Resolution limit	Low contrast sensitivity (1.6 mm detail)	Score
Ref.	1.26	12.5	1.2%	29.23
C–	0.96	11.1	1.6%	26.00
R–	1.24	6.3	0.83%	28.56
N+	1.45	11.1	2.0%	26.88

Table 2. Results of the scores produced by the different test object.

Film ref.	Objective methods		Pseudo-objective methods							
	IQI[6,7]	NIJ[8]	CD-M[8]	FER[9]	NRT[10]	TMAX[11]	TMAM[11]	MTM[12]	NEWC[13]	ACR[14]
C+	0.172	0.741	2.70	11.5	7.40	32.13	78.30	32	38.50	29.5
Ref.	0.179	0.766	2.90	9	6.40	29.23	68.15	40	38.75	27.0
C–	0.209	0.659	3.85	6	4.20	26.00	48.90	13	30.50	12.0
R–	0.207	0.702	2.97	8	4.27	28.56	70.00	22	33.75	23.0
N+	0.214	0.746	3.71	4	5.27	26.88	43.30	13	31.75	10.3
acceptable	<0.195	$>$*?	≤ 3.50	≥ 8	$>$*?	$>$*?	≥ 70.0	≥ 28	$>$*?	≥ 21.0

*? no limit established

Table 3. Relative variations of the scores for the different test object and scoring methods.

Films	Parameters	IQI	NIJ	CD-M	FER	NRT	TMAX	TMAM	MTM	NEWC	ACR
Ref./C–	Contrast	++	++	+++	+++	+++	+	+++	++++	+++	++++
Ref./R–	Resolution	++	+	+	++	+++	+	+	+++	++	++
Ref./N+–	Noise	++	+	+++	++++	++	++	+++	++++	++	++++

CONCLUSION

Definition of reference doses has to be associated with image quality. In mammography several test objects allow image quality to be quantified by giving a single number. To adopt such a procedure, the assessment of the sensitivity of the methodology is of prime necessity. Several pseudo-objective methods have demonstrated a high level of sensitivity, but unfortunately these methods are observer dependent. One fully objective method (i.e. IQI) produced acceptable results; however, some improvements are still needed to increase its sensitivity.

ACKNOWLEDGEMENTS

This work was supported by The Swiss Federal Office for Science and Education (OFES project 94.0037). The authors gratefully acknowledge Dr H. Schibilla from the Commission of the European Communities for the integration of this study in the 'Nuclear Safety Fission' third work programme. This work has been made possible by the participation and the stimulating discussions of the following groups: Pr R. Van Loon, Dr H. Mol and Ms I. De Luyck from the University Hospital of Brussels; M. J. Thurston and Mrs S. House from St George's Hospital of London, Dr D. Dance and Dr E. Castellano from the Royal Marsden Hospital of London; Dr K. Faulkner from the Newcastle General Hospital from Newcastle upon Tyne; Dr A. Noël and J. Stines from the University Hospital of Nancy; Dr M. A. O. Thijssen and Prof. J. H. C. L. Hendricks from the University Hospital of Nijmegen; Prof. J. L. Lamarque, Dr Cherif Cheikh, Ms L. Raynaud and M. S. Sancerni from the 'Institut Montpellierain d'Imagerie Medico-Biologique' IMIM of Montpellier; Prof. E. Vañó, Prof. M. Chevallier and Prof. P. Morán from the University of Madrid; Dr A. Ferro de Carvalho from the Department of Radiation Protection DPSR of Lisbon; Dr M. Fiebich from the University Hospital of Münster; Dr M. Casey from St Vincent's Hospital of Dublin; Dr J. B. Olsen from the Norwegian Radium Hospital of Oslo; Dr A. Servomaa and T. Parviainen from the Department of Radiation Protection (STUK) of Helsinki; Prof. M. Friedrich from the Krankenhavis am Urban Hospital of Berlin; Dr B. Hemdal and Dr I. Andersson from the General Hospital of Malmö; Dr B. Axelsson and Dr B. Hansson from the Karolinsky Hospital of Stockholm; Ms A. T. Klang from the Sahlgrenska University Hospital of Gothenburg; Dr N. Baadegaard from the National Board of Health of Copenhagen; Dr M. Princivalli from the Mestre Hospital (I); Dr R. Gallini from the Unità socio sanitaria locale 41 of Brescia; Prof. Di Maggio and Dr Gambaccini from the University of Ferrara; Dr G. Zonca from the Istituto Nazionale per lo Studio e la Cura dei Tumori of Milan; and Dr A. Flioni-Vyza from St Savas Hospital of Athens.

REFERENCES

1. ICRP. *1990 Recommendations of the International Commission on Radiological Protection*. ICRP Publication 60, Ann. ICRP **21**(1–3) (Oxford: Pergamon) (1990).
2. Warren-Forward, H. M. and Millar, J. S. *Optimization of Radiographic Technique for Chest Radiography*. Br. J. Radiol. **68**, 1221–1229 (1995).
3. Law, J. *Variations in Individual Radiation Dose in a Breast Screening Programme and Consequences for the Balance between Associated Risk and Benefit*. Br. J. Radiol. **66**, 691–698 (1993).
4. Law, J. *Risk and Benefit Associated with Radiation Dose in Breast Screening Programmes — an Update*, Br. J. Radiol. **68**, 870 – 876 (1995).
5. CEC. *European Guidelines for Quality Assurance in Mammography Screening*. 2nd edn, CE-96–96–659-EN-C, (Luxembourg: EC Publications) Luxembourg (1996).
6. Hessler, C., Depeursinge, C, Grecescu, M., Pochon, Y., Raimondi, S. and Valley, J.-F. *Objective Assessment of Mammography Systems*. Radiology **156**, 215–219 and 221–225 (1985).
7. Desponds, L., Depeursinge, C., Grecescu, M., Hessler, C., Samiri, A. and Valley, J.-F. *Image Quality Index for Screen-Film Mammography*. Phys. Med. Biol. **36** 19–33 (1991).
8. Thijssen, M. A. O. Private communication (June 1995).
9. Gambaccini, M., Marziani, M., Cristaudo, L. and De Guglielmo, E. *Performance Device for Quality Control in Mammography*. Radiat. Prot. Dosim. **43**(1/4), 279–281 (1992).
10. Microdata Medicinkonsult AB. *NRT Mammographic Test Phantom* (PO Box 1178, S-171 23 Solna, Sweden).
11. Cowen, A. R., Coleman, J. and Workman, A. *Leeds Mammographic Test Object — Instruction Manual*. FAXIL (The University of Leeds) (1991).
12. Noël, A., Stines, J. and Zerhouni, M. *Développement d'un Fantôme en Matériau Équivalent-sein pour la Mammographie*. Rev. Eur. Biotechnol. Méd. **15**(7), 421–423 (1993).
13. Thompson, S. R. and Faulkner, K. *Phantom for the Measurement of Contrast Detail Performance in Film-screen Mammography*. Br. J. Radiol. **64**, 1049–1055 (1991).
14. Barnes, G. T. and Hendrick, R. E. *Mammography Accreditation and Equipment Performance*. Radiographics **14**(1), 129–138 (1994).