PERGAMON

Applied Radiation and Isotopes 50 (1999) 21–38

Applied Radiation and Isotopes

Measurement of image quality in diagnostic radiology

C.J. Martin[a,*], P.F. Sharp[b], D.G. Sutton[c]

[a]*Health Physics Division, Department of Clinical Physics and Bio-Engineering, West Glasgow Hospitals University NHS Trust, 22 Western Court, 100 University Place, Glasgow G12 8SQ, Scotland, U.K.*

[b]*Department of Bio-Medical Physics and Bio-Engineering, Aberdeen Royal Hospitals NHS Trust, Foresterhill, Aberdeen AB25 2ZD, Scotland, U.K.*

[c]*Department of Medical Physics, Ninewells Hospital, Ninewells, Dundee DD1 9SY, Scotland, U.K.*

Abstract

The aim in radiology is to obtain images which are adequate for the clinical purpose with the minimum radiation dose to the patient. If optimum performance is to be achieved, assessments of image quality must be made to balance against patient dose. The subjective nature of image interpretation makes an objective approach to such assessment difficult. Methods widely applied involve the use of test objects, which although providing a measure of imaging performance may be difficult to link to clinical image formation. The ideal method for evaluation of imaging techniques is through clinical trials and this should be used to address major questions. Scoring of quality criteria, relating to features observed in a normal clinical radiograph, provides a simple method through which image quality can be assessed in every hospital department.

1. Introduction

Medical images should provide sufficient information to allow clinicians to make medical decisions with a reasonable degree of certainty. Better quality images can be obtained in many cases by use of techniques which give higher radiation doses to patients (Martin et al., 1998). Thus in choosing the level of image quality required, the potential risk from loss of diagnostic information from application of a lower dose technique must be weighed against the increased risk that would result from use of a higher radiation exposure. The availability of digital technology offers new opportunities for flexibility in diagnostic imaging. Images can be obtained with significantly lower radiation doses (Marshall et al., 1994; Wade et al., 1995), but considerable care must be taken to ensure that such techniques provide all the clinical data that is required.

It is not possible to monitor accurately all the subtle differences in an X-ray examination, which contribute to a diagnosis. Thus in order to achieve the correct balance between image quality and radiation dose, techniques for evaluating image quality should be set up and links established between results of these tests and clinical performance. Assessments of image quality are particularly important if the best use is to be made of digital systems. Minimum imaging standards should be developed against which comparisons can be made to determine whether equipment performance is adequate for different types of examination. There is a wide variety of approaches to the assessment of image quality and this article reviews some of those available. These include techniques which might be used by equipment evaluation laboratories, equipment manufacturers, hospital physicists, radiologists and radiographers.

2. Adequate image quality for diagnosis

The aim in optimising image quality is to provide an image which is adequate for the clinical task with the minimum radiation dose to the patient. But before a realistic assessment of image quality can be made, the

* To whom all correspondence should be addressed.

0969-8043/98/$19.00
PII: S0969-8043(98)00022-0

requirements need to be defined. It is whether the clinical information required is contained in the image and can be interpreted by the observer that is important, rather than whether the appearance of the image is pleasing to the eye. The ideal set of parameters describing image quality should give a measure of the effectiveness with which an image can be used for its intended purpose, namely answering the clinical questions posed. They should therefore relate to the ability of the image to demonstrate disease and to delineate anatomical structures which are relevant to detection, differential diagnosis and localisation. The image quality requirements vary for different types of examination or even different tasks within single examinations. A low dose option on a fluoroscopic unit will provide a level of image quality which may be entirely satisfactory for some procedures such as barium meals, but unacceptable for others such as angiography. Categories of image quality and radiation dose standards therefore need to be established, which are representative of the requirements for different groups of clinical applications. These should provide standards which can be used to judge whether the performance of imaging equipment is adequate for particular tasks.

Where there is a trade-off between patient dose and image quality, it is necessary for both quantities to be measured. Methods for measurement of patient dose are comparatively well established (IPSM, 1992; Faulkner et al., 1998). Assessment of image quality is less straight forward. Image quality is affected by resolution, sensitivity and statistical noise. Techniques for measuring image quality can be divided into those evaluating images of test objects and those assessing clinical images. Image quality measures derived from the former can be quantified by direct observation. Such evaluations play an essential role in quality assurance programmes because of their simplicity. A major limitation of these techniques is their reliance upon subjective assessment of the image. The observer is making a subjective interpretation of the image, which depends on the ability to identify features relevant to clinical diagnosis (signal) from amid a background of other features and image noise. Image quality can be assessed analytically in terms of signal to noise ratios for different spatial frequencies, which should relate to the ease with which structures of different sizes can be picked out from amid background noise and these methods may be applied to give a more objective assessment of imaging performance (ICRU, 1996).

Measurements using test objects alone describe the behaviour of an imaging system under specific conditions, but it is often difficult to link these directly to clinical performance. Imaging of a patient can be simulated by incorporating a test object in a tissue equivalent phantom, but assessments of clinical images should also be performed to ensure their adequacy for diagnosis. A full evaluation of image performance can be obtained from trials to determine clinical outcomes of cases examined, but such lengthy procedures can only be employed to answer major imaging questions, they are impractical for day to day use. Moreover, the pace of change in technology means that new developments become available before trials on older systems are complete. Diagnostic performance criteria referring to visualisation of anatomical features in normal individuals are being established for quick assessments of clinical images (CEC, 1996a). The techniques that are available for evaluating images of test objects and for assessing clinical images are summarised in Table 1. Methods using test objects placed at the image receptor evaluate performance under ideal conditions, and are not affected by factors such as radiation quality and scattered radiation, while those employing a phantom or clinical image reflect performance in clinical use. Although most techniques will provide a quantitative assessment, most rely on a subjective interpretation of images by observers. The different approaches all have roles to play in the development of optimal image display techniques, in routine monitoring of performance and in judging whether image quality is adequate for particular applications. Measurement of image quality, the range of techniques that are available and the role of each in the overall scheme will be reviewed in the following sections.

3. Use of test objects for assessment of imaging devices

3.1. Measurements of image receptor performance

Measurements of image quality are performed routinely with test objects placed as close as possible to the image receptor, i.e. the image intensifier face or the film cassette (IPEM, 1995, 1996). The techniques attempt to measure the threshold strength at which a signal can be seen in an image. An assessment of resolution can be obtained from a line pair test object containing groups of metal strips with a variety of widths and spacings (Fig. 1(d)) and a measure of threshold contrast from an array of discs of varying contrast (Fig. 1(b)). The "contrast detail diagram" (Fig. 2) is a plot of the minimum detectable contrast for an image feature as a function of its diameter. Such diagrams have usually been measured with test patterns such as the "Rose–Burger" phantom (Burger, 1949, 1950; Rose, 1948, 1973) and the FAXIL test objects (Hay et al., 1985), in which simple signals, such as squares or circles, are present in regular arrays of different sizes with contrast varying regularly in one direction (Fig. 1(a)). The observer must indicate the lowest contrast signal of each size that is visible.

Table 1
Methods for evaluating image quality for X-ray equipment and techniques showing the technical factors taken into account and the type of assessment obtained

Code	Object imaged	Position of object	Data assessed	Method of assessment	Technical factors taken into account	Evaluation quantitative/qualitative	Evaluation subjective/objective
1	test object	image receptor	displayed image	visual scoring	photon fluence (radiation quality)	quantitative	subjective
2			raw image data	digital analysis		quantitative	objective
3			displayed data	digital analysis		quantitative	objective
4	test object in phantom	patient position	displayed image	visual scoring	photon fluence, radiation quality, scattered radiation	quantitative	subjective
5	anthropomorphic phantom	patient position	displayed image	visual assessment	photon fluence, radiation quality, scattered radiation	qualitative	subjective
6	clinical images	patient position	displayed image	visual assessment	photon fluence, radiation quality, scattered radiation	qualitative	subjective
7			displayed image	quality criteria		quantitative	subjective
8			displayed image	clinical trial ROC		quantitative	objective

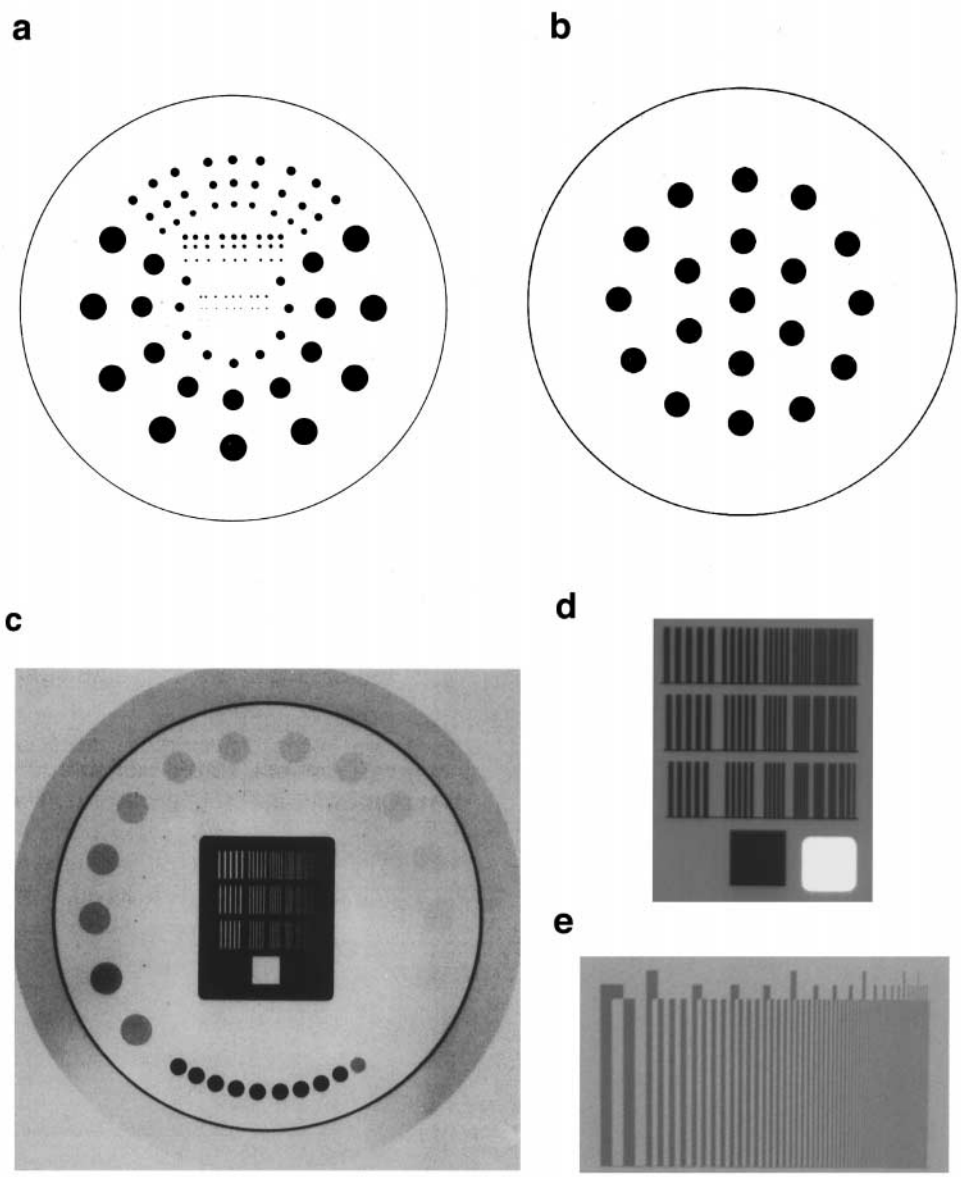

Fig. 1. Diagrams showing the form of FAXiL test objects used for (a) contrast detail and (b) threshold contrast for fluoroscopic units, where all groups of discs of similar size have different contrast levels. (c) An image of the TOR CDR object for film/screen radiography, shown as an inverted image to allow discs of different contrast to be reproduced. (d) the Huttner line pair test object and (e) a line pair test object with groups of lines of different spacing, suitable for use in determinations of MTF.

These tests provide information on imaging capability at different doses or dose rates through visual assessment. They do not provide information on the performance of the system for different radiation qualities or amounts of scattered radiation relating to clinical use, although the system sensitivity may vary with radiation quality. Imaging performance can be compared with levels set down in standards documents (IPEM, 1997), allowing poor performance of an intensifier system to be identified. Similar methods can be used to optimise particular aspects of an imaging system, such as the photon fluence (optical density on the film) required for image perception (Robson et al., 1995), or compare images recorded by different hard copy techniques (Fig. 2(a)). In this example, the film/screen system performs considerably better than the 100 mm camera film, but optimisation of viewing conditions by masking off the illumination in the area sur-

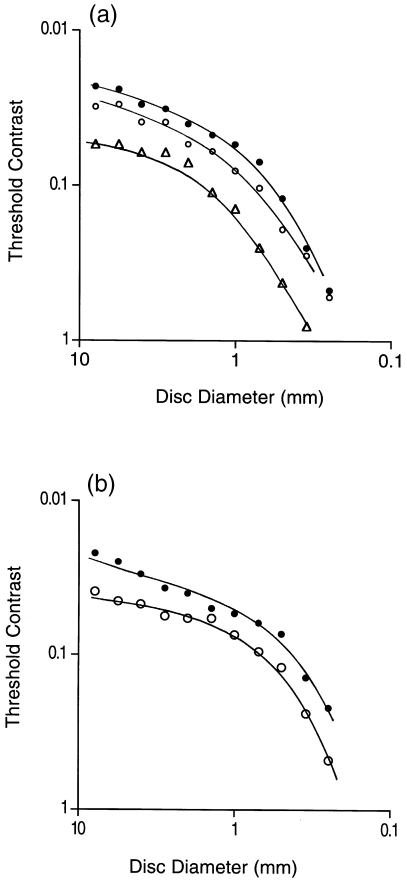

Fig. 2. Contrast detail plots showing differences in detection limits for: (a) images of a FAXIL TO10 test object taken under identical conditions on the same X-ray unit, recorded with a film/screen system (●), a 100 mm camera (△) viewed directly on a light box, and with the same 100 mm film surrounded by a mask to limit stray illumination and viewed with the aid of a magnifier (○): (b) an image of the TO10 test object incorporated into a 100 mm thick perspex phantom to simulate a paediatric examination carried out with (●) and without (○) a grid. Results represent averaged scores for several individuals viewing the films under the same conditions.

rounding the 100 mm film and using a magnifier to examine the image can enhance performance considerably. This demonstrates the importance of standardising viewing conditions for comparisons in quality of different images. The tests rely on a subjective assessment by an observer of whether or not they can detect an object. Since such decisions are affected by both the local conditions and the observer, it is not possible to use the method to make objective comparisons between different fluoroscopic units or make accurate comparisons of equipment performance in different hospitals. Nevertheless, because of the relatively simple nature of the tests, they fulfil an essential role in tasks such as monitoring of image intensifier performance.

For film/screen systems the device output, optical density, is not a linear function of exposure, so the characteristic curve, which relates optical density to exposure (Fig. 3), is an important factor determining imaging performance. This can be determined by exposing the system under test to a range of doses and plotting the variation in density against exposure on a log scale (IPEM, 1996; Hjardemaal and Westergaard, 1992; Warren-Forward, 1995). Important constituents of the test are use of a phantom to simulate modification to the X-ray spectrum by the patient and measurement of the film exposure by an appropriately placed ionisation chamber. Parameters defining the performance of a film/screen system can be determined from the characteristic curve in terms of doses required to give various densities above the base plus fog level (IPEM, 1996). These are the sensitivity or speed, which is the reciprocal of the incident dose to air at the film expressed in mGy to give a density of 1.0, the contrast, which is often expressed in terms of the film gamma defined as the average gradient of the curve between densities of 1.0 and 2.0, and the latitude, which gives the range of exposures that can be accommodated within the normal working density range (0.25–2.0).

3.2. Tests of performance of the whole imaging system

Choice of higher values for tube potential for radiographic or fluoroscopic examinations can often reduce radiation dose to the patient. Test objects which employ metal discs and bars to give contrasting objects cannot be used on their own to assess the impact of changes in radiation quality on clinical images, since

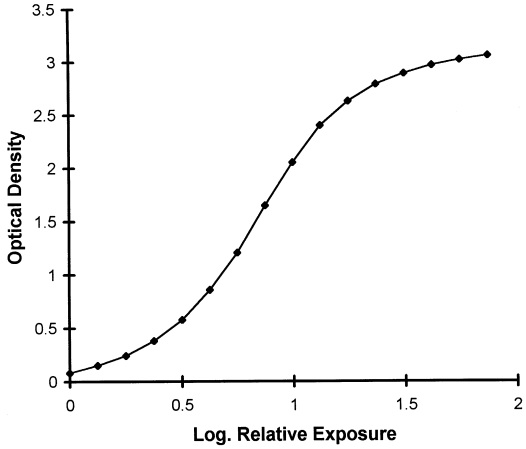

Fig. 3. Characteristic curve of optical density (with base + fog level subtracted) plotted against logarithm of relative exposure for a conventional film/screen system.

attenuation in metals varies with photon energy in a different manner from that in tissue. Moreover, such test objects will not produce significant amounts of scattered radiation, so do not provide information on relative merits of different methods for scatter removal. Phantoms for simulating changes in image quality in clinical examinations should reproduce changes in the quality of the X-ray beam and scattered radiation which occur after it has passed through a standard patient. Assessments have been carried out in which image quality test objects have been combined with uniform attenuating phantoms (ANSI, 1982, Conway et al., 1984, Conway et al., 1990). These phantoms, which are made from plane sheets of perspex and metal, are designed so that the emergent X-ray spectra are similar to those for the body regions they represent. Studies have been carried out with phantoms incorporating contrast detail test objects (Geleijns et al., 1993) and the Leeds TOR(CDR) contrast and resolution test object (Fig. 1(c)) (Llorca et al., 1993; Guibelalde et al., 1994; Vano et al., 1995, Almen et al., 1996) to assess radiographic image quality, and incorporating the Huttner line pair test object (Fig. 1(d)) to assess fluoroscopic and fluorographic systems (Cooney et al., 1995, Nicholson et al., 1995). Both high and low scatter conditions can be simulated by placing the test object nearer to the X-ray tube or the image receptor and results presented in a similar form to those for test objects used on their own (Fig. 2(b)). The sizes of details in the test object are given on the x-axis in Fig. 2(b), but because the object is some distance in front of the film, the image is magnified. The phantoms approximate clinical conditions and good correlation has been observed between such results and ones based on assessments of clinical images (Vano et al., 1995). The phantoms can therefore be used to compare image quality provided by exposure factors selected by different automatic exposure rate control (AERC) options on fluoroscopic equipment.

4. Objective assessment of imaging systems

A more detailed and objective portrayal of the imaging capability of a device can be obtained from a number of physical parameters. Three broad types of parameters have been identified in ICRU (1996).

4.1. The large area (macro) system transfer factor, K

This measures the relationship between the input to the imaging device, X-ray quanta, and the output image, e.g. optical density. For some devices, such as X-ray intensifier tubes, the device output, brightness, is linearly related to the input exposure, the constant of linearity being known as the large-area transfer factor. In other cases, in particular photographic imaging, the detector is non-linear, the large area transfer factor will then be a function of exposure (Fig. 3). In such non-linear systems it is preferable to work in terms of a relative, rather than absolute, scale for both exposure and image brightness and the large scale transfer function is then expressed in terms of the film's gamma.

4.2. Measures of spatial resolution

When considering the response of the imaging system to small features it is necessary to use the detail system response function. This describes the blurring and displacement of the input signals by the imaging system before they form the output. Generally the imaging system is treated as if it is a linear, shift invariant system. The system's response can then be measured in terms of its point, line or edge spread function and the corresponding frequency space representation, the modulation transfer function (MTF), (Fig. 4). A discussion of the measurement of MTF in radiographic systems is given by ICRU (1986). A technique using a bar test object (Fig. 1(e)), for which alignment conditions are less stringent, has also been employed but this involves more approximations (Coltman, 1954; Workman, 1994).

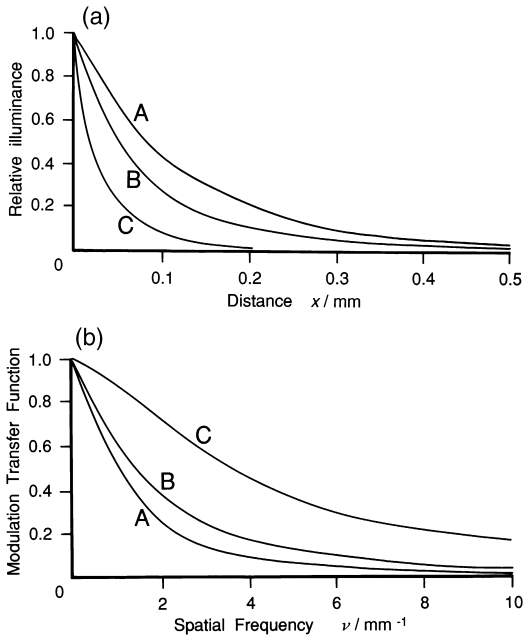

Fig. 4. (a) Line spread functions of (A) a rare earth, (B) a calcium tungstate and (C) a mammographic screen–film system. (b) Modulation transfer functions calculated from the line spread functions (From Doi and Rossman, 1975).

4.3. Measures of noise

Noise may arise from a number of sources; in X-ray film–screen systems it is due to quantum mottle, film granularity, and screen structure mottle (Barnes, 1982). In situations where the signals of interest have low contrast, noise can be regarded as additive, i.e. independent of the signal. Image quality will depend not only on the magnitude of the noise but also its spatial structure. Noise variance analyzed in terms of its spatial frequency content is known as the noise power spectrum, (NPS) or Wiener Spectrum (W_n). The NPS is defined as the Fourier transform of the noise autocovariance function (Fig. 5). Techniques for measuring it can be found in Giger et al. (1984) and ICRU (1996).

While the above parameters have the advantage of being objective, they only measure one particular aspect of device performance and cannot, by themselves, offer a way of judging overall image quality. If noise measured in terms of the output parameters, such as film density, is converted to units of input quantities, such as X-ray photon density, then it is known as the noise equivalent quanta (NEQ). It is given by the equation

$$\mathrm{NEQ}(f) = \frac{K^2 \mathrm{MTF}^2(f)}{W_n(f)}$$

NEQ is the number of quanta that the image is worth based on the three image performance measurements discussed above. If the image was made with Q exposure quanta, then the ratio NEQ/Q is called the detective quantum efficiency (DQE). (Fig. 6).

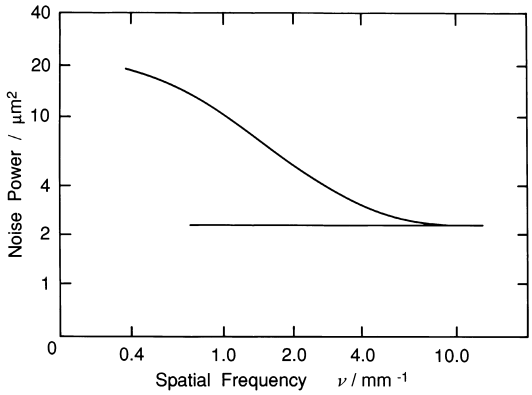

Fig. 5. Noise power spectrum for a film–screen system exposed to X-rays, upper curve, showing quantum mottle and for the film alone exposed to light, lower curve, showing film granularity (From Wagner, 1977).

NEQ can also be linked to the performance of the imaging task, where it summarises the contribution of the imaging system hardware to the performance of the task by the so-called ideal observer. A discussion of ideal and non-ideal observers is beyond the scope of this paper but interested readers can find details and further references in ICRU (1996). Both NEQ and DQE have great potential for measuring equipment performance (see for example Cowan and Workman, 1992), but are not easy to measure in a hospital environment. It may be possible to incorporate software packages into digital imaging devices, which could be

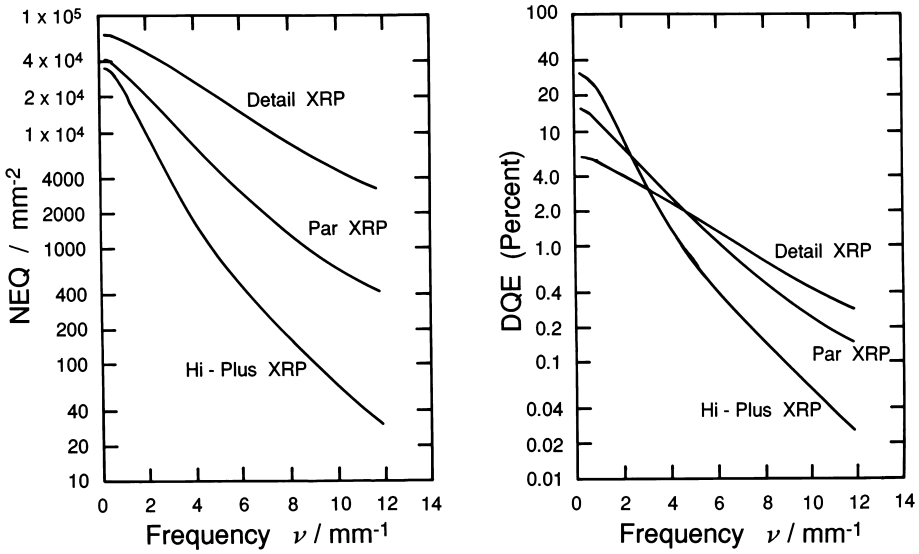

Fig. 6. NEQ and DQE as a function of spatial frequency for three film–screen systems (From Sandrick and Wagner, 1982).

used in assessments of performance for tests carried out under standard conditions in the future.

5. Subjective assessment of imaging systems

Since an image is, by definition, a means for visually representing the clinical information captured by the X-ray equipment, at some stage quality must be based on the judgement of a human observer. Yet much effort has been expended to avoid subjective analysis of image quality. The major problem is perceived to be the lack of reproducibility that appears to accompany such an exercise; firstly large differences in performance have been found between different observers. Secondly, for reasons given below, it is often necessary to view a large number of images to obtain accurate results; it is difficult to persuade observers to spend the time required, and fatigue can, of course, influence performance. In any event, the assessment will be time consuming.

Two general approaches have been taken to subjective analysis; the first is appropriate when the perceptual task requires a binary response from the observer, for example classifying the image as either normal or abnormal, or when comparing a pair of images produced under two different imaging conditions. This allows quality to be measured in terms of specificity and sensitivity. The drawback is that this approach needs the "answer" to be known so that the observer's response can be marked as either true or false. It also requires that the pathology be of borderline visibility. To acquire such a set of clinical images is frequently difficult and the approach is often limited to test patterns where the perceptibility of the abnormal features is under the control of the experimenter. It is, however, a very powerful technique and widely used. It will be looked at in more detail later.

The second approach is to look at relative measures of quality. For example, a set of images produced at different doses on the same machine is presented to the observer who is asked to rank them in order of quality. The definition of quality is under the control of the experimenter and may vary from being as specific as the clarity with which breast calcifications are shown, or as general as simply asking for the images to be arranged in the order of preference (Sharp et al., 1982). By repeating the exercise with either different sets of images and/or different observers, quality can be judged by the agreement in ranking order. While the strength of agreement will give an indication of the superiority, in this case, of the images produced at one dose level compared with another, it does not permit a quantitative judgement to be made as to how much better the images produced at one dose level is than another. Nor, of course, does it answer the question as to whether a lower dose may still give images of sufficient quality for the task. A more quantitative assessment can be made by assigning a scale value to the quality of images or specific features within the image.

Much of the work on subjective assessment of image quality has been based on the first approach. This includes the methods such as contrast detail plots (Fig. 2) described in the section on use of test objects for assessment of imaging devices. This type of test pattern has the great advantage of being very simple to use, but the cost is the poor statistical reliability of the results produced. There is usually only a single signal for each contrast and size combination, yet for signals close to the limits of perceptibility, noise effects result in quite significant random fluctuations in visibility. The method of constant stimulus (MCS) attempts to take this effect into account in its definition of threshold visibility. The technique requires a test object in which signal contrast can be varied by the experimenter. The observer is shown a set of images containing the signal at contrasts varying from that which is sufficiently high for the signal to be clearly visible, to low values at which the signal cannot be seen. For each selected contrast value, several images are produced. The number of occasions on which the feature at a particular contrast level is seen gives the true positive response; a plot of true positive against contrast gives a visual response curve (Fig. 7). Typically the threshold contrast is taken to be that which produces a 50% visual response.

Both approaches have the advantage of giving a measure of image quality in a physically realistic form; it is possible to link the contrast and size to the features likely to be seen in clinical practice. While the

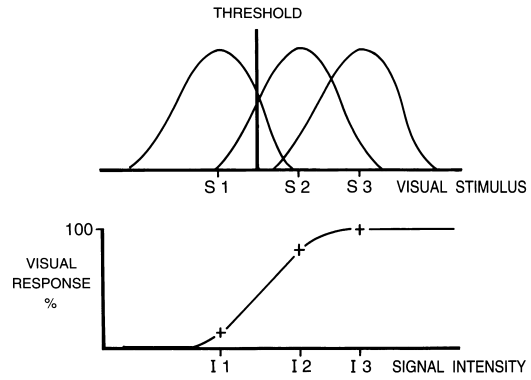

Fig. 7. The signals of intensity $i1$ to I3 will produce visual stimuli of average intensities $s1$ to $s3$. It is assumed that the observer adopts a visual threshold T such that a signal is seen in the image only if the stimulus that it produces exceeds this visual threshold. The resulting curve showing true positive response rate as a function of signal intensity is known as the visual response curve.

method of constant stimulus gives a more reliable measure of threshold detectability, both it and the contrast–detail approach suffer from one fundamental problem, they assume that the observer's visual system has some intrinsic perceptual threshold that must be exceeded by the visual stimulus produced by the test pattern before the signal can be seen. This threshold theory of vision was challenged in the early 1950s and has been supplanted by statistical decision theory. One consequence is the realisation that an observer can consciously vary his/her visual threshold and hence the assessed quality of the image.

6. Signal detection theory

While the observer's decision making process is represented by an internal threshold which must be exceeded by the visual stimulus produced by the signal for the signal to be seen, this threshold can be consciously varied by the observer according to the degree of confidence the observer requires in his/her judgement before reporting the signal as seen. A high degree of confidence will imply that a high visual stimulus, i.e. a high contrast signal, is needed before the observer reports the signal as seen, while if the observer employs a low confidence level, then signals with much lower contrast are seen. Thus a technique is needed which provides a measure of image quality which avoids the influence of this internal decision criterion.

Consider the situation in which the observer is presented with a set of images some of which are normal and some contain an abnormality, the task being to differentiate between the two. Not only may the observer correctly identify either an abnormal image (true positive response) or normal image (true negative), but they may also incorrectly identify either a normal image as abnormal (a false positive response) or an abnormal image as normal (a false negative response). The situation is represented in Fig. 8. As the decision criterion/threshold is varied so will the number of true and false responses. As the threshold increases, the number of true positive responses and the number of false positives will decrease. A plot of true against false positives yields a curve, such as shown in Fig. 9, known as a receiver operating characteristic (ROC) curve (Green and Swets, 1966; Metz, 1978, 1986a; Swets and Pickett, 1982). The curve shows the effect of changes in decision criterion. If the image quality changes such that it becomes easier to differentiate normal from abnormal images, then the position of the curve will also vary (Fig. 10).

Thus the position of the curve is a measure of the discriminability between normal and abnormal images, the closer it gets to the upper left hand corner of the graph the better the image quality, while points making up each curve reflect the changes introduced by the observer varying the visual threshold. Thus the ROC curve provides a way of separating out the distinguishability of signal from noise from the observer's ability to alter the visual threshold.

While the relative position of ROC curves gives a ranking of the quality of image sets, a quantitative measure of image quality is still desirable. A number of such measures are available, the area under the ROC curve being perhaps the most widely used. While such measures measure the discriminability between signal and noise, they do not provide the physically

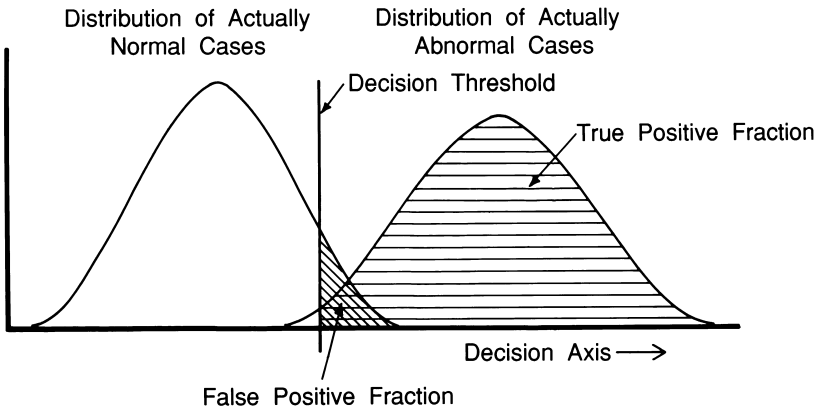

Fig. 8. The visual stimuli produced by the set of normal and abnormal images are represented by Gaussian curves, reflecting the variation in image appearance. The observer adopts a threshold which depend upon his/her decision criteria. Images producing a visual stimulus greater than the threshold value are reported to be abnormal. Hence the fraction of the abnormal image curve exceeding the threshold is the true positive fraction and that of the normal image curve the false positive fraction.

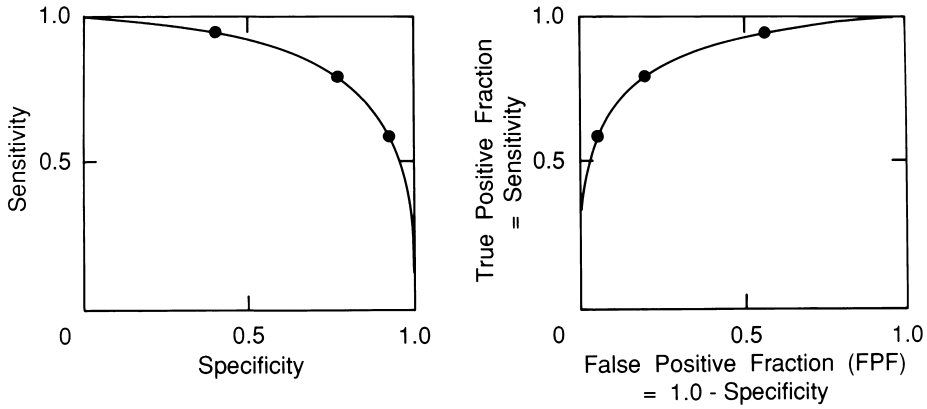

Fig. 9. The receiver operating characteristic (ROC) curve, which can be plotted as either sensitivity against specificity or true positive response rate against false positives.

meaningful measure of image quality found with the method of constant stimulus and contrast detail curves.

The most common and most efficient way of generating ROC curves is by using the rating technique (Green and Swets, 1966; Goodenough et al., 1974; Metz, 1979; Swets and Pickett, 1982). The observer is asked to assess the likelihood of the image being abnormal using a scale such as shown in Table 2. Alternatively a continuous scale has been suggested (Rockette et al., 1992). A number of tests have been proposed for measuring the statistical difference between ROC curves, the principal ones being McNeil and Hanley (1984), Hanley (1989), Metz (1986b, 1989) and Dorfman et al. (1992). The application of signal detection theory to radiology was reviewed by Metz in 1986. Recently it has been used to study the effect on diagnostic accuracy of reducing patient dose (Roehrig et al., 1997).

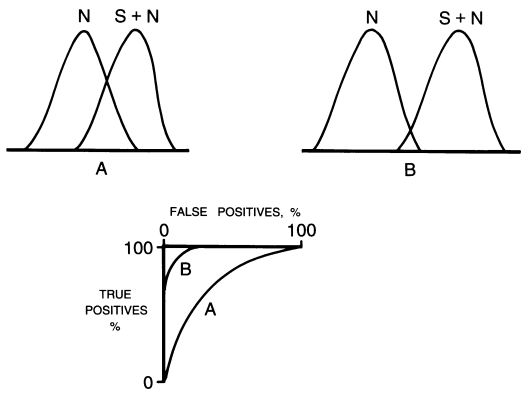

Fig. 10. As the image quality improves then the overlap between the curve representing the visual stimulus produced by images containing a signal (S + N) and those containing no signal (N), decreases, i.e. one moves from the situation shown in A to that in B. This results in the ROC curve (curve A) shifting further towards the top left-hand corner of the plot of true positives against false positives (curve B).

6.1. Relationship between ROC and MCS

The method of constant stimulus approach and signal detection theory are not independent ways of measuring quality, but two "views" of the same set of data (Sharp, 1990). The surface shown in Fig. 11 illustrates how the surface cut in one direction yields the visual response curve, and in the other the ROC curve.

6.2. Forced choice experiments

As the name suggests, this approach forces the observer to make a decision. In the simplest form the observer is faced with two images, only one of which contains a signal, i.e. is abnormal. The observer's task is to identify which image contains the signal. Thus the problem of a varying internal threshold is overcome by requiring only relative responses from the observer. This is known as the two-alternative forced choice experiment (Green and Swets, 1966). The number of correct responses is linked to the area under the ROC

Table 2
Rating scale

Rating	Description
1	abnormality almost certainly not present
2	abnormality probably not present
3	abnormality possibly present
4	abnormality almost certainly present
5	abnormality definitely present

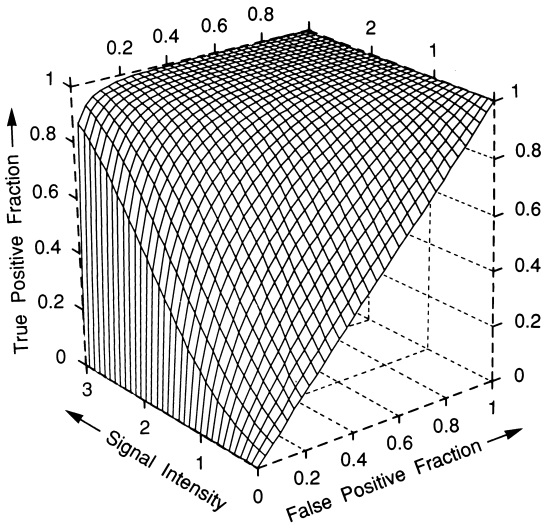

Fig. 11. A three-dimensional graph showing the relationship between visual response curves and ROC curves. The surface can be considered to be made from a set of visual response curves, showing true positive fraction as a function of signal intensity, at different false positive values. Alternatively, it is made from a set of ROC curves, true against false positive fractions, for a series of signal intensity values.

curve (Green and Swets, 1966; Hanley and McNeil, 1982). A multiple alternative forced choice experiment involves more than two images, but still only one contains a signal. The advantage of the forced choice approach is that the experiments are more reproducible and the results have an unambiguous interpretation (ICRU, 1996). Forced choice experiments have also been used to produce contrast–detail diagrams (Loo et al., 1984).

7. Assessment of clinical images

7.1. Clinical trials

At one end of the spectrum of measures relating to image quality are those giving objective measures of specific, technical aspects of imaging system performance, such as MTF for spatial resolution. NEQ then provides a way of combining these measures into a single unit of performance. To give a measure of image quality, however, NEQ needs to be combined with information about the specific task required of the observer; the ideal and non-ideal observers provide a framework for doing this, yielding a signal to noise ratio (ICRU, 1996). While these models link the physical parameters into a decision model, they do not themselves provide a simple mathematical alternative to the human observer, at the best giving a measure of ideal performance.

Fryback and colleagues (Fryback and Thornbury, 1991) have proposed a six-level efficacy model summarised in Table 3. ROC based studies, which sit at the next level of efficacy, provide a technique for measuring the quality of the image by a human observer in a way that permits the removal of bias associated with the confidence level adopted by the observer in making his/her decision. For simple tasks, the quality measures derived from ROC curves can be reconciled with those from the ideal and non-ideal observers, see ICRU (1996). Since quality measures are task-dependent, then ultimately, the effectiveness of imaging systems require to be evaluated against real clinical data. Trials employing real clinical data are usually difficult to perform and time consuming; this is a major problem in circumstances where the technology of the imaging system may well change significantly during the time of the trial. A first step, therefore, is to analyze the clinical task to see if it can be modelled by a simple task which can be incorporated in a test pattern. To date, it should be noted, quality measures have often relied upon well defined but simple detection tasks; these are classification tasks in which the response required refers to a discrete number of categories, often just two. Work needs to be done to extend the methodology to estimation tasks which involve the measurement of a continuous parameter, such as the spatial extent or intensity of the signal.

While the role of objective measures of quality cannot, at present, utilise real clinical data, no such

Table 3
Six-level model of efficacy

Level	Typical output measures
Technical efficacy	MTF
Diagnostic accuracy efficacy	sensitivity, specificity; ROC curve analysis
Diagnostic thinking efficacy	change in clinician's diagnostic probability
Therapeutic efficacy	percentage of times therapy changed
Patient outcome efficacy	change in quality adjusted life years
Societal efficacy	summed quality adjusted years; Positive change in national product

restriction applies to the ROC approach. Indeed it carries the important message that the defining of clinical performance in terms of single values of sensitivity and specificity is not only oversimplified, but may also be inaccurate. The ROC curve shows the trade-off between sensitivity and specificity for a specific set of imaging conditions. The particular combination that is most efficacious depends both upon the prevalence of the disease in question in the population studied and the benefits and costs of the various correct and incorrect decisions in the diagnostic setting (McNeil et al., 1975; Metz et al., 1975; Metz, 1978; Swets and Swets, 1979; Swets and Pickett, 1982; Sainfort, 1991). By considering these the combination of true and false positive responses on the ROC curve that yields the highest benefit can be determined. This benefit can then be compared with the "cost" of performing the procedure. This provides the methodology by which quality can be considered in the context of the higher levels of efficacy.

8. Assessment of clinical images in a radiology department

8.1. Image quality criteria

The ROC approach provides methodology which can be applied to clinical studies. Simpler techniques are required for carrying out routine assessments in X-ray departments for evaluating local performance and for deciding whether techniques are appropriate for different applications. Although the important imaging requirement is detection of abnormalities, most assessments of clinical image quality in a radiology department must be based on visualisation of normal anatomy since the majority of images are normal, and technical assessment of a range of abnormalities would be impractical. Guidelines on Quality Criteria for clinical radiographs have been developed by the Commission of the European Communities (CEC) which set out diagnostic requirements for a radiograph of a normal adult for common examinations (CEC, 1996a). The relevance, acceptability and ease of use of the criteria have been assessed in trials involving 83 departments in 16 countries (Maccia et al., 1995), as well as in studies on large numbers of radiographs in single departments (Vano et al., 1995). These criteria include a list of anatomical structures, which should be reproduced and other features which should appear visually sharp (Table 4). They encompass positioning of the patient as well as the imaging capability of the system. Typically about 90% of the image quality criteria are fulfilled by radiographs regarded as acceptable for clinical diagnosis.

Table 4
Diagnostic requirements for a Chest PA projection (CEC, 1996a)

	Image criteria	Positional technique/ image quality
1.		
1.1	performed at full inspiration and with suspended respiration	position
1.2	symmetrical reproduction of the thorax as shown by central position of the spinous process between the medial ends of the clavicles	position
1.3	medial border of the scapulae to be outside the lung fields	position
1.4	reproduction of the whole rib cage above the diaphragm	position
1.5	visually sharp reproduction of the vascular pattern in the whole lung, particularly the peripheral vessels	image quality
1.6	visually sharp reproduction of: the trachea and proximal bronchi, the borders of the heart and aorta, the diaphragm and lateral costo-phrenic angles	image quality
1.7	visualization of the retrocardiac lung and mediastinum	image quality
1.8	visualization of the spine through the heart shadow	image quality
2.	Important image details	
2.1	small round details in the whole lung, including the retrocardiac areas: high contrast: 0.7 mm diameter, low contrast: 2 mm diameter	image quality
2.2	linear and reticular details out to the lung periphery: high contrast: 0.3 mm in width, low contrast: 2 mm in width	image quality

The methodology provides a valuable practical tool for assessment of clinical image quality, but has limitations. It is based on visualisation of normal anatomy for which the requirements may be less demanding than for differentiation of abnormal pathologies. It does not take into account variations in observer thresholds in defining the limit of acceptability and significant variation in responses of different radiologists have been reported, with field radiologists tending to score films as acceptable more often than "experts" (Maccia et al., 1995). In addition, scores may be influenced if patients with abnormal pathology are included as this may mask the visibility of some features (Wraith et al., 1995, McParland et al., 1996). Despite these drawbacks, the clinical image quality criteria provide the best method that is available for assessing the quality of radiographs in individual departments and mark an important step in the standardisation of techniques. For local studies where images obtained using different techniques are being compared, criteria grouped under one heading can be separated, if that is considered appropriate, or a half mark awarded for fulfilment of each.

In addition to the image quality criteria, the guidelines include an indication of the size for important image details which the system should be capable of visualising (Table 4), examples of equipment and exposure factors considered representative of good radiographic technique and reference values for patient entrance surface doses for an adult patient. Departments should be able to work within the reference doses without compromising image quality. If the mean dose for any examination within a department exceeds the reference dose, this should be used as a trigger for action to optimise the technique. A similar methodology has been applied to paediatric radiography (CEC, 1996b) and CT examinations (CEC, 1997). Many aspects, particularly those relating to radiation doses for CT, still need to be established, but the documents represent an important first step in the development of standards. The reference doses do not put limits on lowest doses and recent surveys have shown that doses significantly less than the guidelines can be achieved. However, it is essential that effects on image quality are taken into account when making any changes to reduce dose.

The area of radiological imaging where clinical assessment is more difficult is fluoroscopy. Videos or sets of digital images can be recorded for evaluation by several radiologists (Smiddy et al., 1996), but image quality criteria are more difficult to apply, so that a qualitative assessment of an examination may be the best that can realistically be employed. In addition, different dose rate options may be used for various parts of an examination depending on the image quality requirements at each stage. Because of these difficulties, assessment of image quality by the use of test phantoms is the method generally adopted for fluoroscopic systems. However, more guidelines on suitable tube potentials and image performance requirements for fluoroscopic and fluorographic images recorded from image intensifiers for different types of examination would be valuable.

8.2. Practical assessment of clinical image quality

The clinical radiographic quality criteria (CEC, 1996a) are designed to be applied in audits of image quality in an X-ray department. Some criteria include visualisation of several elements and these should be scored independently. Assessments should be undertaken for a random selection of patients of average size (60–80 kg) and equipment and exposure factors should be recorded for each radiograph. Entrance surface doses should be evaluated and methods for carrying this out are described elsewhere in this issue (Faulkner et al., 1998; Jessen et al., 1998; Geise, 1998; Dance and Skinner, 1998). Forms listing the image criteria, which should be fulfilled, must be completed by at least two radiologists observing each radiograph and the scores averaged. Each observer should work independently of the others. Studies must include sufficient numbers of images that the results are not unduly influenced by one or two patients. This would normally be a minimum of ten. Any consistent non-compliance with image quality criteria should then be investigated and corrective action taken where this is considered necessary.

When carrying out assessments, care should be taken to standardise viewing conditions by ensuring that all light boxes used are of the same type, illumination colour and brightness. Before starting the trial, some experience should be gained in observing films and filling in assessment forms, as observer thresholds have been seen to change during the early stages of evaluations of this type. Having a set of "ideal" films in which all aspects of performance have been optimised can be helpful. Such films must have been taken with exposures below the corresponding reference level. The test films can then be compared directly with the ideal ones to aid in assessment. The paediatric quality criteria (CEC, 1996b) also include a system for scoring more general aspects of an image such as film blackening, contrast and sharpness. These should not be used in deriving the image quality score, but may aid in interpretation of results.

The method can be employed to assess changes in technique. An initial evaluation can be made using an anthropomorphic phantom (ICRU, 1992; White, 1993). However, when a technique is introduced clinically, a comparison of radiographs taken with the old and new techniques should be made to ensure that the

new one represents a real improvement and that any radiation dose saving does not compromise the diagnostic potential of the image. Only criteria relating to the reproduction of clinical features should be used for such a comparison. When selecting images for inclusion, attention should be paid to obtaining groups of patients with a similar mix of age and a similar mean weight to represent each technique. Any patients for whom the disease process is likely to influence the result should, as far as possible, be excluded. A sample of at least ten should be used for both the original technique and the new one, and this should be increased to twenty or more if there is a likelihood of the pathology of individual patients having a significant influence on the results. The images should be judged by at least two radiologists working in the relevant area and preferably more. None of those carrying out the assessment must be involved in the selection of patients. All markings indicating the source of each image should be masked, where practicable, and images randomised to eliminate bias from position in the assessment. It is only through evaluation of clinical images in addition to the image performance tests carried out on X-ray equipment that a department can ensure that all aspects of its performance have been optimised and that the standard of performance is maintained.

Establishment of the process of audit for both dose and image quality is one of the most important steps for improving performance in an imaging department. Assessments encourage knowledge of technical aspects of equipment performance and how these relate to image quality and patient dose. Significant improvements can often be made in radiation dose and image quality where performance is closely monitored and such improvements maintained. A proper dose/image quality control audit programme is particularly important for systems which have a broad exposure latitude. Studies of chest radiographs performed on mobile equipment with a broad latitude film/screen combination have revealed both a wide range of exposures and a significantly higher mean exposure than was used for a combination with a narrower latitude, but similar speed index (Simpson et al., 1998). When a quality control audit of both radiation dose and image quality was carried out and results fed back to users a substantial reduction in radiation dose was achieved with no measurable effect on image quality. Similar findings have been reported with computed radiography systems which can provide satisfactory images with an even wider range of exposures (Seibert et al., 1996). Image quality/dose audit and feed back should play an important role in achieving optimal performance for all digital systems.

Table 5
Techniques for evaluating image quality showing the part of the imaging system tested and the application. (The code refers to that used in Table 1)

Code	Object imaged	Position of object	Method of assessment	Part of imaging system tested	Task performed	Group using technique
1	test object	image receptor	visual scoring	imaging system and display	monitoring performance	local
2			digital analysis	system excluding display	system evaluation	evaluation group
3			digital analysis	imaging system and display	system evaluation	evaluation group
4	test object in phantom	patient position	visual scoring	system as in clinical use	comparison of techniques	local
5	anthrop. phantom	patient position	visual assessment	system as in clinical use	preliminary assessment of clinical technique	local
6	clinical images	patient position	visual assessment	system as in clinical use	clinical technique evaluation	local
7			quality criteria	system as in clinical use	clinical technique evaluation	local
8			clinical trial ROC	system as in clinical use	clinical technique evaluation	international assessment

9. Discussion

There are many ways in which images can be assessed and the factors influencing the various methods are given in Table 1. These methods all have roles to play in evaluation of new systems, routine monitoring of performance, and the comparison and evaluation of clinical techniques, and these are summarised in Table 5. Methodology for assessment of image quality is still being developed, but it is important that techniques already available are applied in regular evaluations for individual departments in order to improve and maintain imaging performance. This is particularly important for digital systems which can provide acceptable images for a wide range of doses, because of their broad dynamic range.

The simplest method for evaluating the performance of imaging equipment is to use an arrangement with a test object placed as close as possible to the image receptor. This avoids many of the variables present in clinical use, which may degrade the image, so that a direct assessment can be made of true performance under ideal conditions. Evaluations of this type are simple to carry out and are employed routinely in hospitals with observers grading images of a variety of test objects to provide a guide to performance and check for any deterioration. However, the tests rely on observations by individuals who will have differences in threshold criteria for making judgements and so do not provide objective data suitable for evaluating and comparing different systems. Parameters such as the NEQ and DQE have potential for providing objective assessments of imaging performance for an image as viewed by an ideal observer. Measurements of this type should allow objective comparisons to be made between different systems. The methods are not easy to apply in a hospital environment, but can be employed by equipment manufacturers and evaluation groups for assessing the performance of individual models. More research is required to establish links between these measures of performance and requirements for clinical examinations. In addition, standards of performance relating to different groups of clinical examinations need to be set which can be used to judge whether the image quality provided by a particular unit is adequate for the intended clinical purpose.

The methods discussed in the preceding paragraph evaluate imaging performance under ideal conditions. They do not take account of how differences in radiation quality would affect the image of a real patient or how scattered radiation will degrade the image. These factors can only be included through assessment of clinical images or images of phantoms with similar properties to those of a patient. Test objects can be incorporated into tissue equivalent phantoms and scored in terms of the features that can be visualised. This method may be of particular value for applications such as comparing the imaging performance of fluoroscopic equipment at different AERC exposure factor options (Martin et al., 1998), and choosing the optimum exposure factor programme for different examinations. However, the final test of imaging performance must be of the system in clinical use. Image quality criteria which should be fulfilled by normal clinical images provide a valuable guide to imaging performance which can be used in hospital departments. The method can be employed both for assessing the overall quality of clinical images in a department and for evaluating changes in technique to ensure that they provide an improvement in performance.

In order for the optimisation of imaging systems to proceed further, there is a need for standards to be set for imaging performance and for radiation doses relating to a range of different radiological examinations. The only method that can establish objectively the suitability of a technique is a clinical trial. ROC analysis methods can be applied to remove the bias associated with the confidence level adopted by the observer and so provide objective assessments. These methods require a considerable amount of time and effort. However, they can and should be applied to answer major clinical imaging questions, such as the determination of optimum imaging conditions and minimum acceptable doses for a range of common examinations. Finding answers to these questions is particularly important if digital techniques are to be used to their full potential. Much research in this area has involved individual hospitals working in relative isolation and there is now a need for a coordinated approach. Careful consideration must be given to the imaging parameters radiation quality, photon fluence and removal of scattered radiation, which affect both image quality and radiation dose (Martin et al., 1998), and conditions standardised in participating departments. Information gained from these studies should enable optimum performance criteria to be established for a range of clinical applications in radiology for both conventional and digital imaging.

10. Conclusions

A range of techniques are available for assessing image quality. Involvement of X-ray equipment manufacturers, medical physicists, radiologists, radiographers and others in the application of these techniques at various stages in equipment development, installation and use should enable the gradual improvement in imaging performance to be continued. Further developments are required, including setting of standards against which imaging performance can be judged through the organisation of large scale clinical

trials and application of objective methods for measuring performance. Major studies of this type should be the next step towards achievement of the optimal balance between patient dose and image quality.

References

Almen, A., Loof, M., Mattson, S., 1996. Examination technique, image quality and patient dose in paediatric radiology — a survey including 19 Swedish hospitals. Acta Radiologica 37, 37–342.

American National Standards Institute (ANSI), 1982. Method for sensitometry of medical X-ray screen–film processing systems. ANSI PH2.43. ANSI, New York.

Barnes, G. T., 1982. Radiographic mottle: A comprehensive theory,. Med. Phys. 9, 656–667.

Burger, G. C. E., 1949. The perceptibility of details in Roentgen examinations of the lung,. Acta Radiol. 31, 193–222.

Burger, G. C. E., 1950. Phantom tests with X-rays,. Philips Tech. Rev. 11, 291–298.

Coltman, J. W., 1954. The specification of imaging properties by response to a sine wave input. Journal Optical Society of America 44, 468–471.

Commission of the European Communities (CEC), 1996. European guidelines and quality criteria for diagnostic radiographic images. EUR 16260 EN, ed. J. H. E. Carmichael, C. Maccia, B. M. Moores, J. W. Oestmann, H. Schibilla, D. Teunen, R. Van Tiggelen and B. Wall. CEC, Brussels.

Commission of the European Communities (CEC), 1996. European guidelines on quality criteria for diagnostic radiographic images in paediatrics. EUR 16261, ed. M. M. Kohn, B. M. Moores, H. Schibilla, K. Schneider, H. St. Stender, F. E. Stieve, D. Teunen and B. Wall. CEC, Brussels.

Commission of the European Communities (CEC), 1997. Quality criteria for computed tomography working document. EUR 16262. CEC, Brussels.

Conway, B. J., Butler, P. F., Duff, J. E., Fewell, T. R., Gross, R. E., Jennings, R. J., Koustenis, G. H., McCrohan, J. L., Reuter, F. G., Showalter, C. K., 1984. Beam quality independant attenuation phantom for estimating patient exposure from X-ray automatic exposure controlled chest examinations. Medical Physics 11, 827–832.

Conway, B. J., Duff, J. E., Fewell, T. R., Jennings, R. J., Rothenberg, L. W., Fleischman, R. C., 1990. A patient equivalent phantom for estimating patient exposure from automatic exposure controlled X-ray examinations of the abdomen and lumbo-sacral spine. Medical Physics 17, 448–453.

Cooney, P., Marsh, D. M., Malone, J. F., 1995. Automatic exposure control in fluoroscopic imaging. Radiation Protection Dosimetry 57, 269–272.

Cowan, A. R., Workman, A., 1992. A physical image quality evaluation of a digital spot fluorography system,. Phys. Med. Biol. 37, 325–342.

Dance, D. R. and Skinner, C. L., 1998. Breast dosimetry. Applied Radiation and Isotopes 50, 185–203.

Doi, K. and Rossman, K., 1975. Measurement of optical and noise properties of screen–film systems in radiography, *Proc. Soc. Photo-Optical Instrumentation Engineering*, Vol. 56, pp. 45–53.

Dorfman, D. D., Berbaum, K. S., Metz, C. E., 1992. ROC rating analysis: generalization to the population of readers and cases with the jacknife method,. Invest. Radiol. 27, 723–731.

Faulkner, K., Broadhead, D. A. and Harrison, R. M., 1998. Patient dosimetry measurement methods. Applied Radiation and Isotopes 50, 115–125.

Fryback, D. G., Thornbury, J. R., 1991. The efficacy of diagnostic imaging,. Med. Decis. Making 11, 88–94.

Geise, R. A., 1998. Radiation dose in interventional fluoroscopic procedures. Applied Radiation and Isotopes 50, 173–184.

Geleijns, J., Scultze Kool, L. J., Zoetelief, J., Zweers, D., Broerse, J. J., 1993. Image quality and dosimetric aspects of chest X-ray examinations: Measurements with various types of phantoms. Radiation Protection Dosimetry 49, 83–88.

Giger, M. L., Doi, K., Metz, C. E., 1984. Investigation of basic imaging properties in digital radiography. 2. Noise Wiener spectrum,. Med. Phys. 11, 797–805.

Goodenough, D. J., Rossman, K., Lusted, L. B., 1974. Radiographic applications of receiver operating characteristic (ROC) analysis,. Radiology 110, 89–95.

Green, D. M. and Swets, J. A., 1966. Signal Detection Theory and Psychophysics. Wiley, New York. Reprinted with corrections and an updated topical bibliography by Krieger, Huntington, NY, 1974 and by Peninsula Publishing, Los Altos, CA, 1988.

Guibelalde, E., Fernandez, J. M., Llorca, A. and Ruiz, M. J., 1994. Image quality and patient dose for different screen-film combinations. British Journal of Radiology, 67, 166–173.

Hanley, J. A., 1989. Receiver operating characteristic (ROC) methodology: the state of the art,. Crit. Revs. Diag. Imag. 29, 307–335.

Hanley, J. A., McNeil, B. J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143, 29–36.

Hay, G. A., Clark, O. F., Coleman, N. J., Cowen, A. R., 1985. A set of X-ray test objects for quality control in television fluoroscopy. British Journal of Radiology 58, 335–344.

Hjardemaal, O., Westergaard, H., 1992. The speed of medical X-ray screen/film processing systems in Danish Hospitals. Radiation Protection Dosimetry 42, 115–119.

International Commission on Radiation Units and Measurements (ICRU), 1986. Modulation Transfer Function of Screen–Film Systems, ICRU Report 41. International Commission on Radiation Units and Measurements, Bethesda, MD.

International Commission on Radiation Units and Measurement (ICRU), 1992. Phantoms and computational models in therapy diagnosis and protection. ICRU Report 48, Bethesda, MD.

International Commission on Radiation Units and Measurements (ICRU, 1996. Medical Imaging — the assessment of image quality, ICRU Report 54.

International Commission on Radiation Units and Measurements, Bethesda, MD.

Institute of Physics and Engineering in Medicine (IPEM), 1995. Measurement of the performance characteristics of diagnostic X-ray systems used in medicine. X-ray image intensifier television systems, ed. P. A. Hiles and H. C. Starritt. IPEM Report No. 32, Part II, 2nd edn. Austin, Hertford, U.K.

Institute of Physics and Engineering in Medicine (IPEM), 1996. Measurement of the performance characteristics of diagnostic X-ray systems used in medicine. X-ray intensifying screens, films, processors and automatic exposure control systems, ed. M. R. Holubinka. IPEM Report No. 32, Part IV, 2nd edn. Austin, Hertford, U.K.

Institute of Physics and Engineering in Medicine (IPEM), 1997. Recommended standards for the routine performance testing of diagnostic X-ray imaging systems. IPEM Report No. 77. IPEM, York.

Institute of Physical Sciences in Medicine (IPSM), National Radiological Protection Board (NRPB), College of Radiographers (CoR), 1992. National Protocol for patient dose measurements in diagnostic radiology. NPRB, Chilton.

Jessen, K. A., Shrimpton, P. C., Geleijns, J., Panzer, W. and Tosi, G., 1998. Dosimetry for optimisation of patient protection in computed tomography. Applied Radiation and Isotopes 50, 165–172.

Llorca, A. L., Guibelalde, E., Varno, E., Ruiz, M. J., 1993. Analysis of image quality parameters using a combination of ANSI type phantom and the Leeds TOR (CDR) test object in simulations of simple examinations. Radiation Protection Dosimetry 49, 47–49.

Loo, L.-N. D., Doi, K., Metz, C. E., 1984. A comparison of image quality indices and observer performance in the radiographic detection of nylon beads,. Phys. Med. Biol. 29, 837–856.

Maccia, C., Ariche-Cohen, M., Nadau, X., Severo, C., 1995. The 1991 CEC trial on quality criteria for diagnostic radiographic images. Radiation Protection Dosimetry 57, 111–117.

McNeil, B. J., Keeler, E., Adelstein, S. J., 1975. Primer on certain elements of medical decision making,. New Engl. J. Med. 293, 211–215.

McNeil, B. J., Hanley, J. A., 1984. Statistical approaches to the analysis of receiver operating characteristic (ROC) curves. Med. Decis. Making 4, 137–150.

McParland, B. J., Gorka, W., Lee, R., Lewall, D. B., Omojola, M. F., 1996. Radiology in the neonatal intensive care unit: dose reduction and image quality. British Journal of Radiology 69, 929–937.

Marshall, N. W., Faulkner, K., Busch, H. P., Marsh, D. M., Pfenning, H., 1994. A comparison of radiation dose in examination of the abdomen using different radiological techniques. British Journal of Radiology 67, 478–484.

Martin, C. J., Sutton, D. G. and Sharp. P. F., 1998. Balancing patient dose and image quality. Applied Radiation and Isotopes 50, 3–21.

Metz, C. E., 1978. Basic principles of ROC analysis,. Seminars Nucl. Med. 8, 283–298.

Metz, C. E., 1979. Applications of ROC analysis in diagnostic image evaluation. In *The Physics of Medical Imaging: Recording System Measurements and Techniques*, ed. A. G. Haus, p. 546. Am. Inst. Phys., New York.

Metz, C. E., 1986a. ROC methodology in radiologic imaging,. Invest. Radiol. 21, 720–733.

Metz, C. E., 1986. Statistical analysis of ROC data in evaluating diagnostic performance. In *Multiple Regression Analysis: Applications in the Health Sciences*, ed. D. Herbert and R. Myers, p. 365. Am. Inst. Phys., New York.

Metz, C. E., 1989. Some practical issues of experimental design and data analysis in radiological ROC studies,. Invest. Radiol. 24, 234–245.

Metz, C. E., Starr, S. J., Lusted, L. B., and Rossmann, K., 1975. Progress in evaluation of human observer visual detection performance using the ROC curve approach. In *Information Processing in Scintigraphy*, ed. C. Raynaud and A. E. Todd-Pokropek, p. 420. CEA, Orsay, France.

Nicholson, R. A., Thornton, A., Akpan, M., 1995. Radiation dose reduction in paediatric fluoroscopy using added filtration. British Journal of Radiology 68, 296–300.

Robson, K. J., Kotre, C. J., Faulkner, K., 1995. The use of a contrast–detail test object in the optimization of optical density in mammography. British Journal of Radiology 68, 277–282.

Rockette, H. E., Gur, D., Metz, C. E., 1992. The use of continuous and discrete confidence judgements in receiver operating confidence judgements in receiver operating characteristic studies of diagnostic imaging techniques,. Invest. Radiol. 27, 169–172.

Roehrig, H., Krupinski, E. A., Hulett, R., 1997. Reduction of patient exposure in pediatric radiology,. Acad. Radiol. 4, 547–557.

Rose, A., 1948. The sensitivity performance of the human eye on an absolute scale. J. Opt. Soc. Am. 38, 196–208.

Rose, A., 1973. Vision: Human and Electronic. Plenum Press, New York.

Sainfort, F., 1991. Evaluation of medical technologies: a generalized ROC aproach.. Med. Decis. Making. 11, 208–220.

Sandrick, J. M., Wagner, R. F., 1982. Absolute measure of physical image quality: Measurement and application to radiographic magnification. Medical Physics 9, 540–549.

Seibert, J. A., Shelton, D. K., Moore, E. H., 1996. Computed radiography X-ray exposure trends. Academic Radiology 3, 313–318.

Sharp, P. F., 1990. Quantifying image quality,. Clin. Phys. Physiol. Meast. 11, 21–26 Suppl. A.

Sharp, P. F., Chesser, R. B., Mallard, J. R., 1982. The influence of picture element size on the quality of clinical radionuclide images. Phys.Med.Biol. 27, 913–926.

Smiddy, P. F., Quinn, A. D., Freyne, P. J., Marsh, D., Murphy, J. M., 1996. Dose reduction in double contrast barium enema by use of low fluoroscopic currents. British Journal of Radiology 69, 852–854.

Simpson, P. D., Martin, C. J., Darragh, C. L. and Abel, R., 1998. A study of chest radiography with mobile X-ray units. *British Journal of Radiology*, 71, in press.

Swets, J. A. and Pickett, R. M., 1982. *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory*. Academic Press, New York.

Swets, J. A. and Swets, J. B., 1979. ROC approach to cost/benefit analysis. In *Proceedings of the Sixth IEEE*

Conference on Computer Applications in Radiology, p. 203. IEEE Computer Society Press, New York.

Vano, E., Guibelalde, E., Morillo, A., Alvarez-Pedrosa, C. S., Fernandez, J. M., 1995. Evaluation of the European image quality criteria for chest examinations. British Journal of Radiology 68, 1349–1455.

Wade, J. P., Goldstone, K. E., Dendy, P. P., 1995. Patient dose measurement and dose reduction in East Anglia (U.K.). Radiation Protection Dosimetry 57, 445–448.

Wagner, R. F., 1977. Fast Fourier digital quantum mottle analysis with application to rare earth intensifying screen systems. Medical Physics 4, 157–162.

Warren-Forward, H. M., 1995. What is the true film–screen sensitivity? Radiation Protection Dosimetry 57, 301–304.

White, D. R., 1993. The specifications of 25 phantoms presented at the Wurzburg Workshop (June, 1992). Radiation Protection Dosimetry 49, 371–386.

Workman, A., 1994. Physical evaluation of the imaging performance of Dupont Ultra-Vision screen–film systems. Medical Devices Agency Report 94/40, FAXIL, Leeds, U.K.

Wraith, C. M., Martin, C. J., Stockdale, E. J. N., McDonald, S., Farquhar, B., 1995. An investigation into techniques for reducing doses from neonatal radiographic examinations. British Journal of Radiology 68, 1074–1082.