NEW TECHNOLOGY IN RADIOLOGICAL DIAGNOSIS: AN INVESTIGATION OF DIAGNOSTIC IMAGE QUALITY IN DIGITAL DISPLAYS OF RADIOGRAPHS

Submitted for the Degree of Doctor of Medicine
in the University of London
by

RICHARD M DAWOOD, BSc MB BS FRCR

JANUARY 1991

ABSTRACT

Digital radiology is undergoing rapid evolution. Its objectives can be summarized as the creation within the modern radiology department - and indeed within the entire hospital - of a harmonious, integrated, electronic network capable of handling all diagnostic radiological images, obviating the need for conventional film-based radiology.

One of the limiting factors in the introduction and exploitation of digital technology is the issue of image display quality: if electronic display systems are to be widely used for primary radiological diagnosis, it is essential that the diagnostic quality of the displayed images should not be compromised. From the perspective of the practising radiologist, this study examines the performance of the first two commercially available digital radiological display systems to be purchased and installed in a British hospital.

This work incorporates an extensive observer performance investigation of image quality from existing 1024- and 1280- line display systems, and suggests that displayed images digitized at a pixel size of 210μm show a significant reduction in diagnostic performance when compared with original film. Such systems appear to be unsuitable for primary radiological diagnosis of subtle lesions.

Some of the physical properties of such systems, some relevant methodological issues, and the relationship between image quality and other factors influencing the development acceptance and implementation of digital technology, have also been investigated; the results are presented.

This is a controversial subject, and conflicting views have been expressed in the British literature concerning the issue of whether or not the technology is now ready for total system implementation; the view of this author is that careful testing of display systems, and of every other component of digital networks, should precede their entry into clinical use.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

The following radiologists kindly participated in one or more of the observation experiments:

A Al-Kutoubi, FRCR; C J Barber, FRCR; A H Choudhri, FRCR; J O M C Craig, PRCR; M E Crofton, FRCR; D A Cunningham, FRCR; J H Highman, FRCR; R W Kerslake, FRCR; M McCarty, FRCR; P Rowlands, FRCR; J M Stevens, FRCR.

I am grateful to them for their stamina, perseverance and forbearance.

Dr H M Saxton kindly permitted the use of case material from Guy's Hospital to be included in the renal osteodystrophy series.

Dr Alison Maclean kindly acted as an external reviewer of the mammography series.

David Plummer of the Department of Medical Physics, University College London, examined data downloaded from the digitizers to enable measurement of signal to noise ratio.

Ms Dawn Beaumont, of the Computer Unit at St Mary's Hospital, provided invaluable assistance in processing the results.

Ms Jane Wadsworth, of the Academic Department of Community Medicine, St Mary's Hospital Medical School, provided advice

and practical assistance with statistical methods and interpretation.

The Department of Audiovisual Services at St Mary's provided much assistance throughout the duration of the project, and has assisted with preparation of the Figures.

I am grateful to Professor Harold Kundel for his helpful comments and suggestions on a draft manuscript of this work.

Ms Alison Porter was the research radiographer responsible for the operation of the display systems installed at St Mary's; her expertise and professionalism contributed greatly to this work, and I wish to acknowledge her advice and encouragement throughout the duration of this project.

I am also indebted to Andrew Todd-Pokropek, Reader in Medical Physics at University College London, who has been a generous source of expertise and good counsel at every stage, and who kindly agreed to supervise my preparation of this thesis.

This work was conducted in the Department of Radiology at St Mary's Hospital, and further experience arose from visits and discussions at numerous research sites abroad. I should like to record my particular gratitude to Dr Oscar Craig and my other colleagues at St Mary's, without whom this work could not have taken place, and also to the many colleagues in the UK and abroad who have contributed to ideas put forward in this thesis.

1. INTRODUCTION

1.1. THE PROBLEM

New technological developments in medical imaging are making it possible for radiologists and clinicians to transact exclusively in digital images across electronic networks, supplanting film-based imaging systems. The expectation is that radiological images will in future be displayed on some form of electronic display system for diagnosis.

There are now solutions for many of the early technical impediments to such innovation: one of the most important issues for which there is still no clear answer, however, is the question of diagnostic image quality. Will such images hold sufficient diagnostic information to accomplish their purpose? Radiographic images have been acquired on film for almost as long as X-rays have been used in medical diagnosis. The effects upon radiological diagnosis of displaying these images electronically rather than on film are largely unknown; an essential requirement for the new technology is that the diagnostic information content of the displayed images, and the ability of the radiologist to perceive and utilize that information, should remain unimpaired. The diagnostic implications must be studied in detail before such new technology can be introduced.

1.2. AIMS

The present study comprises a series of experiments intended to determine whether or not a significant loss of diagnostic image quality attends the use of current, commercially available digital radiological display systems, using genuine clinical case material. The systems to be considered are the first two commercially available display systems to be installed in the UK.

Specifically, it has been widely supposed that 1000 x 1000 matrix ("1K") display systems are entirely suitable for primary radiological diagnosis, and the investigation is intended to explore this hypothesis.

Certain physical factors emerged in the course of the investigation as being important elements in the context of diagnostic image quality; some of these will be considered in detail. It also became necessary to examine the validity of the methodology that was being employed.

Image quality is not the only parameter of importance in relation to overall clinical diagnostic performance - for example, the time taken to retrieve and display images. Such issues are important in relation to implementation of the new technology, and are also examined.

These issues will be placed in the context of the recent history, current status, and likely future developments in digital radiology.

1.3. BACKGROUND

The late Seventies and early Eighties saw unprecedented advances in the technology of medical diagnostic imaging. Ultrasound, CT scanning, new gamma cameras and new radio-isotope imaging technology, low-osmolar contrast media, digital subtraction angiography, and magnetic resonance imaging, all followed in rapid succession. These developments brought new power to the modern radiodiagnostic armoury; because of them, radiology is now able to make substantial contributions to diagnosis and management in almost every branch of medicine. This was a period that also saw sharp increases in the price of silver, and consequently in the cost of film. The microchip revolution was in full swing, and computer systems of every type were achieving new standards of performance at lower cost.

The sophistication of the new imaging methods, however, contrasted sharply with our ability to manage the resulting images, which had essentially remained static. An internal audit of X-ray films taken in the Accident and Emergency Department of a major London teaching hospital (Dawood, 1983) showed that between 20 and 30 per cent of films had

been lost or mis-placed within the first month following examination, and could not be recovered despite an exhaustive search; more than 45 per cent of films belonging to patients admitted as emergencies had still not been seen by a radiologist one month later. Accurate data relating to this type of administrative failure is not readily available, and is costly and time-consuming to collect. Many institutions do not gather such information, or are understandably reluctant to publish their findings; steps have recently been taken to formalize medical audit as part of a package of reforms to the National Health Service (Department of Health, 1989) so it is possible that sources of data may improve. Experience indicates, however, that the figures quoted above are comparable with those from other departments both in the UK and the USA: loss rate figures for the University of Kansas are quoted at 18%, despite computerization and the introduction of bar coding (S Dwyer, personal communication), and national figures for Japan at 20% (Mun and Akisada, 1989).

All radiologists are familiar with the frustrations of working with film-based images in a large hospital setting, and the problems relate not just to lost or mis-filed films and lost or mis-matched reports. As demand increases, there is an ever-growing problem of providing radiologists and referring clinicians alike with exactly the images they need, when and where they need them. Hard pressed junior medical staff devote much of their time to tracking down films and reports on their patients for clinical

presentations and case conferences; for one 530-bed hospital it has been estimated that such tasks occupy the efforts of the equivalent of 5 whole-time junior doctors (Glass & Slark 1990). Many of the problems can be traced to poor motivation among the low-paid staff employed to handle and transport the images.

Until 1990, legal requirements in the UK stipulated that films should be retained for at least eight years following radiographic examination, or - in the case of children - until the age of twenty-one; these requirements were justifiable, though not widely practised, even though experience suggests that only 3-5 per cent of patients' film folders that have been stored for one year will ever be required again. Requirements have now been modified to allow individual hospitals to negotiate suitable retention periods with their own health authority, but in other countries, films may have to be retained for even longer periods (e.g. 20 years in Finland). Film is expensive - in 1989, St Mary's Hospital spent roughly £150,000 on film. It is also bulky, heavy, and costly to store.

Use of film requires a quality assurance programme that must be assiduously maintained. Film reject rates in excess of 10 per cent are commonplace in most hospitals, which means further radiation exposure to patients and staff, further delay to other patients, and further wear on equipment. Processing requires chemicals that pollute the environment, are unpleasant to handle, and cause side

effects to staff that are sometimes serious, ranging from skin damage and reactions to inhalational injury, cardiac arrhythmias, vocal cord paralysis, asthma, allergies and sensitization (Gordon 1989). Film recycling and silver recovery are possible, but are inefficient, time-consuming, and expensive; revenue from silver recovery does not always return to the X-ray department.

1.4. DIGITAL RADIOLOGY: EVOLUTION OF A NEW CONCEPT

The concept of a digital network began to evolve as a possible solution. The earliest reported instances of video handling, display, communication or storage of radiographic images had in fact taken place in the USA in the late Sixties, but the concept of an integrated network is largely attributed to Paul Capp, at the University of Arizona. It was, and is, appealing in its apparent simplicity. Many of the images produced in an X-ray department are already electronic: ultrasound images, CT scans, radionuclide scans, digital subtraction angiography and magnetic resonance scans account for up to a third of the examinations currently undertaken in modern hospital radiology departments. They yield digital data, and give measurements of physical parameters from which diagnostic images are constructed. These images are traditionally printed out on film, but they do not need to be. Why not create an electronic network within the department - throughout the hospital, even - that would make every image instantly accessible wherever and whenever it is needed.

The development of fully integrated picture archiving and communication systems, or PACS, as such technology came to be known, seemed an attractive solution.

Conventional radiography accounts for the remaining seventy per cent of the workload. For the PACS ideal to be accomplished, it is necessary to put those images into a digital form as well. A number of early systems were proposed to replace conventional radiography, using specialized equipment with an X-ray source, some kind of scanning process, and an array of electronic detectors (Fraser et al, 1983; 1987; 1989); or a CT scanner (Huebener 1983). They were expensive, bulky, unwieldy, and unsuitable for use on sick people. They also did not take account of the huge investment in conventional, existing radiographic equipment. It is worth bearing in mind that in Britain, the current NHS replacement cycle for radiographic equipment can be as long as thirty years. What is more, the images they yielded were of poor quality. Attempts have also been made to derive static radiographs from digitization of output from image intensification systems.

One of the most crucial technological advances to date in image acquisition has been Computed Radiography (CR), which was developed in 1981 (Takano 1982; Hachiya et al 1982; Sonoda et al 1983; Takahashi et al 1984; Tateno 1984; Tateno et al, 1987). This replaces conventional X-ray film with a re-usable radiation-sensitive phosphor storage

medium. Following exposure using conventional radiographic equipment, the sheet of phosphor is removed from its cassette and undergoes a high-intensity laser scanning process that stimulates luminescence of the stored analogue image, enabling it to be converted into digital data. Initially developed by Kodak and manufactured in Japan by Fuji, several companies now make or market such systems. CR is more sensitive than film and has much wider exposure latitude: experience shows that images of consistent appearance can be readily obtained under a wide range of exposure conditions, even under difficult circumstances such as in theatre or in the intensive care unit (Kangarloo et al 1988). Some reports have suggested that radiation doses can be considerably reduced - by up to 98 per cent or more in certain situations (e.g. follow-up spinal views for scoliosis, Kogutt 1987; pelvimetry, Aoki et al, 1987) and reductions of 30-50 per cent have been claimed; and in theory anyway, the need for chemical processing and film could be eliminated. For the first time, it realizes the prospect of being able to enhance, manipulate and interrogate projection radiographs for improved diagnosis. Other potential applications of CR are also worthy of note: they include portal radiography (in radiotherapy); and electrophoretic autoradiography, in which increases in sensitivity of up to 100-fold have been reported (Johnston et al, 1990).

In its purest form, the PACS ideal allows linkage of all of these technologies into a computerized network that could

permit radiologists and clinicians to transact exclusively in digital images on display systems within the X-ray department, throughout a hospital and between hospitals for consultation.

Further historical aspects are considered in the Literature Review (page 34).

1.5. PERCEIVED BENEFITS AND MOTIVATION

Diagnostic radiology is by far the largest source of artificial exposure of the public to radiation. There is a widespread expectation that digital radiology would allow significant reductions from this source to be made, by factors ranging from 30-98 per cent.

The high film loss rate in most hospitals means that examinations are often repeated because original films cannot be found, with attendant unnecessary further risk and radiation exposure. Many people expect that electronic image handling will make a dramatic difference to patient management and care, and to considerably enhance our ability to provide the best possible service. Some authors have concluded that this would also lead to shorter stays in hospital (e.g. Glass & Slark, 1990).

The use of digital images would enable radiologists to manipulate, electronically process, and interrogate images for improved diagnosis; it would also facilitate

communication of images within and between hospitals, enabling better sharing of expertise and improved communication, and would also open up new areas of research.

Numerous cost analyses have been conducted, and some of these will be referred to in the Literature Review. The task is complex because there are many possible variables. There is confidence in many circles that cost savings from reduced film usage make a powerful cost-justification for the new technology.

Further perceived benefits include the following:

- images instantly available for viewing throughout hospital at multiple locations simultaneously

- no need for film storage, filing systems, porterage etc

- no more reject films from incorrect exposure (though half of all rejects are from positioning errors, however)

- longer lifespan of equipment from use of lower exposure factors

- integration of images with hospital/departmental computer systems

- multimodality processing, viewing and interaction

- ability to communicate images to other institutions for conferences and expert opinion

The quest for alternatives to film-based systems is strongly motivated, and the pressure for rapid development and implementation of PACS technology continues to grow.

1.6. IMAGE QUALITY: EMERGENCE AS A MAJOR ISSUE IN PACS

Crucial to any possible success of such technology is the ability to handle and display images without impairing their clinical diagnostic value. High resolution video monitors have been developed for this purpose. They are expensive, and inevitably represent a compromise between a host of competing technical factors. For example, as the number of lines on the TV system increases, bandwidth must also increase or flicker will occur, which would adversely affect the ability of observers to make a diagnosis.

There are many technical difficulties associated with PACS, and these will be discussed in due course. In the UK, early recognition of the role of display quality as a crucial element in the viability of digital radiological systems is to the credit of the team working at St Mary's led by Dr Oscar Craig. This group has had the opportunity to study displayed image quality on two commercially available display systems, firstly with 1000-line monitors,

and subsequently with 1280-line monitors, both using digitized film. Results of these investigations form the basis of the work presented here.

1.7. DIGITAL RADIOLOGY IN THE UK: St MARY'S HOSPITAL

The first hospital in the UK to consider a large-scale implementation of digital radiology was St Mary's Hospital, London.

St Mary's Hospital was much in need of redevelopment. Its X-ray department was a grim and gloomy place (Figure 1.1), scattered in a haphazard manner through the basement of a hospital that pre-dated Röntgen by half a century. Despite this, it had been a centre of excellence for at least since the post-war years, and these surroundings added impetus to the motivation to make the new department as modern as possible.

Plans were drawn up in 1979, with a new wing expected to open in 1987, housing a new department of diagnostic radiology conceived initially along conventional lines. Detailed consideration of the equipment to be installed began in 1982, just as awareness and interest in new imaging technology had reached its peak. The PACS concept was entirely new, and the arrival of all the technology necessary to implement what seemed initially to be a very simple concept, was perceived to be just around the corner.

This appeared to be an ideal opportunity to explore the introduction of PACS technology. A formal feasibility study was commissioned by the Department of Health in 1982, and this reported in 1985 that the objectives were just possible. The new St Mary's was therefore planned as a totally filmless hospital, and the timing seemed perfect to implement the PACS ideal in its entirety. It was expected that there might be a few technical problems, just as with any new technology, but with such rapid progress on every other radiological frontier it seemed that they would soon be resolved. Linking all the equipment together, and linking the radiology department up to the rest of the hospital, appeared to be a relatively simple matter.

By the end of 1985, however, it had become clear that the necessary technology was simply not available. Funding in excess of 9 million pounds had been promised, and had not been the problem; but no manufacturer could supply what was required, with the necessary guarantees that a clinical service could actually be provided - it should not be forgotten what the technology is intended to achieve.

It remains true that no hospital is currently "filmless", anywhere in the world. The new wing at St Mary's (Figure 1.2) nonetheless housed a limited digital display system that became the focal point of digital research and evaluation in the UK.

Figure 1.1: The old radiology department at St Mary's

Figure 1.2: The new wing at St Mary's (1987)

In a technology that had so far largely been the preserve of computer scientists, physicists and engineers, and to a considerable extent remains so, the St Mary's project was an opportunity to give prominence to the requirements of radiologists, investigating methods of measuring and comparing image quality, and drawing attention to the need for objective, scientific evaluation of each component of a PACS network. The project focused upon diagnosis, and its impact has been surprisingly great considering the discrepancy in funding compared with PACS projects in other countries.

From the perspective of the practising radiologist, the objective was to evaluate commercially available display systems in terms of their clinical radiological diagnostic performance. An inescapable issue, as far as diagnostic radiologists and their patients are concerned, is the question of whether or not we will be able to make the correct diagnosis.

2. LITERATURE REVIEW

2.1. INTRODUCTION

Problems with storage, retrieval and display of valuable images are not new. They were recognized, for example, by the painter Pannini in 1759 (Figure 2.1).

The earliest reported attempts at electronic transmission and storage of radiographic images were made at the University of California, San Francisco, more than two decades ago. Miller and McCurry (1969) proposed setting up closed circuit TV links between the radiology department and the wards; a TV camera would send pictures of the day's radiographs to recorders on each ward, where relevant images would be stored on fourteen-inch aluminium videodiscs, and would be available for viewing on TV screens. The theoretical advantages of such a system became immediately apparent. It appears that a cost justification for such a scheme was produced, but experience seems to have been limited to a single machine, and no further record of the venture is evident.

Steckel (1972), working from University of California, Los Angeles, set up an 875-line closed circuit TV system to transmit images from radiographs in the radiology department to a ward. No details are available of the

Figure 2.1: Gallery of scenes of ancient Rome: Pannini, 1759

technical specifications of the display. A zooming facility was present, and the purpose of the installation was to allow radiologists to conduct daily case conferences and teaching without having to leave their department.

Image quality issues did not arise, since in neither case were electronic images to be used for primary diagnosis; the original film radiographs would be retained - though possibly with greater security since they would not have to leave the radiology department.

The origin of the concept of an integrated electronic network is largely attributed to Paul Capp, working in the University of Arizona. In 1973, the possibilities for a "photoelectronic" department were explored (Capp 1981). The first reference to such systems in the Japanese literature is believed to be by Iinuma (1974). Such systems became known as PACS, an acronym that was formally adopted at the first international meeting on the subject, organized by Samuel Dwyer from the University of Kansas, at Newport Beach, California, in 1982.

At first, it seemed that only small advances in technology would be required. In 1984, Gray felt that the technology was "available or just on the horizon" (though he debated whether it was totally desirable, stating that such technology could only be justified if it would make the users better radiologists and would improve current radiological practice. In 1981, Capp had predicted that by

the year 2000, every radiological department would be digital. In 1985, Capp et al considered that the previous estimate had been too conservative, and wrote: "it now appears that this change will occur at least 5-10 years earlier than predicted"; they reaffirmed that "within 5 to 10 years, radiology departments will most likely be totally electronic, cost-effective and more diagnostically accurate". Forecasts like these enable the theorist to speculate upon the prevailing view regarding the state of maturity of the technology at any point during its development.

Technological advance has been accompanied by increasing recognition of the complexity of the problems. Delegates to IMAC '89, an international conference on electronic image management in Washington DC, were unwilling to contemplate the likelihood of total systems becoming viable before 1995. Fraser et al (1989) nonetheless predict that digital technology will have entirely replaced film-based imaging in half of all teaching hospitals by the year 2000. Schwenker (E Du Pont de Nemours & Co, conference paper, Steamboat Springs 1990) has observed that predictions of the likely timescale for implementation of PACS have been showing a relentless increase ever since the earliest days of the technology, and suggests that the trend will probably continue, in the short term at least.

In the late 1980s, at a time when the first digital departments were expected to have been functioning, the

limitations of digital technology have had to be accepted. Plans for a totally digital department at St Mary's Hospital in the UK had to be shelved due to inadequate performance, operational difficulties and reservations concerning the lack of any quantitative study of image quality, as well as non-availability of suitable hardware from any major supplier (Craig 1988, Dawood 1989). The same considerations have influenced plans in hospitals elsewhere: two of the most ambitious schemes, at the Madigan Army Hospital, Washington, and the Hokkaido University Hospital in Japan, have had to be delayed or radically modified; and even at the University of California at Los Angeles, which has a clear technological lead in this field, the current plan is now for a phased implementation extending over a minimum of a five year period (Huang 1990); by 1993, only one third of the imaging procedures performed at UCLA will be available on the network, and the single factor restricting implementation at this site is considered to be the availability of suitable technology rather than its cost (Kangarloo 1990, conference paper, NATO ASI).

2.2. SPATIAL RESOLUTION REQUIREMENTS

At the heart of the issue of image quality in digital radiology is the question of the spatial resolution requirement for radiological diagnosis. Increasing the spatial resolution of digital systems carries a high

penalty: the higher their resolution, the greater their complexity and cost, and the longer the community will have to wait for their maturation and implementation. Defining the minimum acceptable spatial resolution is a difficult task, and accordingly, relatively few large scale studies have been published: these will now be considered in detail.

One of the earliest studies on diagnosis from displayed images was undertaken by Andrus et al in 1975, in an investigation of the feasibility of teleradiology. The proposal was to establish a microwave transmission system between two hospitals, using a TV camera to capture images of radiographs, and interactive controls to allow the remote radiologist to zoom or scroll around the image. In the study, a mixture of skull, chest and abdominal films were used, 100 in all; the resulting images were displayed on 512-line monitors to five observers. The results proved unsuitable for receiver operating characteristic (ROC) analysis, but the authors considered the results to be of "acceptable accuracy" in view of the participants' inexperience with the new display.

At around the same time, Revesz and Kundel set up a microwave link to hospitals affiliated to Temple University School of Medicine, Philadelphia, to enable transmission of images during hospital "grand rounds". Preliminary studies on pulmonary nodule detection appeared to show equivalent

performance for the transmitted images (Revesz & Kundel, 1973).

In 1981, Foley et al conducted an ROC study of diagnosis from chest images. They used 40 chest radiographs, of which 20 were normal, 10 showed miliary shadowing, and 10 showed multiple small nodules. The diagnosis of normality or otherwise was determined by consensus of an independent panel. A region of interest within the lung field, measuring 10cm x 20cm, was digitized from each film with a scanning microdensitometer with a 0.2mm aperture. Images were then printed out onto hard copy to give pixel sizes of 0.2, 0.4, 0.6, 0.8, 1.0, and 1.6mm (2.5-0.3125 lp/mm). Images with pixel sizes larger than 0.2mm were generated by pixel averaging. These were viewed by 10 observers. Original film images were not included in comparisons. Results showed improved lesion detection with smaller pixel size, but no statistically significant improvement in performance between the 0.2, 0.4, 0.6, 0.8, or 1.0mmm pixel sizes. They concluded that there was no loss of accuracy for nodule detection until pixel size increased above 1 mm, and commented that increasing resolution had resulted in more false positives; a 1mm pixel size might be adequate for radiodiagnosis.

For reference, the relationships between matrix size, pixel size, resolution in line pairs per mm, and data volume, are summarized in Tables 1 and 2. Interestingly, there are no comments in the literature regarding the question of pixel

TABLE 1: Scanning spot size: theoretical relationship to matrix size, bit depth, resolution and image data.

Spot size	Matrix size	Line pairs/mm	Bit depth	Grey levels	Mbytes raw data
200μm	1750 x 2150	2.5	8 bit	256	3.76
			10 bit	1024	4.7
			12 bit	4096	5.64
100μm	3500 x 4300	5	8 bit	256	15.05
			10 bit	1024	18.81
			12 bit	4096	22.575

Note: Values are given for a 35cm x 43cm image (e.g. full-sized chest radiograph) because this is the largest film size used in conventional practice, and corresponds to the area scanned by most current, commercially available laser film digitizers. Matrix sizes given for digitizers usually refer to maximum scan areas, so that the area scanned with smaller films will always be less (for the same resolution). In practice, image files are always larger - on account of the necessary header and format data attached to each image.

TABLE 2: Matrix size: relationship to pixel size, resolution, and image data. Theoretical values for a 35cm x 43cm image (e.g. chest radiograph).

Matrix Size		Pixel size (mm)		Line pairs/mm	No of Pixels	Mbytes (8 Bit)	Mbytes (10 Bit)	Mbytes (12 Bit)
64	64	5.47	6.72	-	4096	0.0041	0.0051	0.0077
128	128	2.73	3.36	-	16384	0.01638	0.0205	0.0307
256	256	1.37	1.68	-	65536	0.06554	0.0819	0.12
512	512	0.68	0.84	0.74	262144	0.26	0.33	0.49
1024	1024	0.34	0.42	1.47	1048576	1.05	1.31	1.97
2048	2048	0.17	0.21	2.94	4194304	4.19	5.24	7.86
4096	4096	0.09	0.10	5.55	16777216	16.78	20.97	31.46
8192	8192	0.04	0.05	12.5	67108864	67.11	83.89	125.83

shape; pixels are in practice neither square nor rectangular, but round or ovoid.

In another widely cited study, Lams and Cocklin (1986) digitized a 20cm x 20cm region of interest from each of 38 chest radiographs at 0.2, 0.4, 0.8, and 1.0mm pixel sizes, again using a scanning microdensitometer. The films showed solitary nodules (12 cases), septal lines (10 cases), or normal lung (16 cases); diagnosis had been agreed by a panel, and the images were viewed by eighteen observers. With the nodules, they concluded that pixel sizes smaller than 0.8mm did not yield statistically significant improvements in performance. However, the requirement to resolve septal lines was 0.4mm. They accepted that the need to include consensus cases in this study might have weighted the clinical material in favour of grosser cases, thereby underestimating true resolution needs, and they called for wider application of ROC methods during introduction of digital technology.

Both of these studies addressed portions of images, under highly artificial viewing conditions. The suggestion that a pixel size of 0.4mm might be adequate - readily achievable with 1000-line monitors - caused a surge of interest.

Carterette et al (1986) studied 100 images of an anthropomorphic chest phantom, of which 50 were "normal" and 50 showed a small simulated nodular lesion. The images

were digitized at a matrix of 2048 x 2048, and displayed on a 1024-line monitor. The images were viewed by 5 observers. They found no statistically significant difference in diagnostic performance between the displayed images and the original film.

MacMahon et al (1986) studied 40 normal and 40 abnormal (interstitial shadowing and pneumothorax) chest radiographs, digitized at 10 bits through an aperture of 0.1mm, and printed onto hard copy to yield images with pixel sizes of 1.0, 0.5, 0.2, 0.1mm. Diagnosis was established by clinical history plus the consensus opinion of two reviewers. A region of interest showing two thirds of a single lung, and excluding the hilum, was shown to twelve observers in order of increasing resolution. Results showed increased accuracy with increasing resolution, but all of the digital images performed significantly worse than the original film.

In one of the first studies on entire images from real patients, Goodman et al (1986) used 150 images digitized at a 1680 x 2000 matrix and 12 bits per pixel. They consisted of 100 abnormal chest radiographs (showing interstitial lines/nodules, pulmonary nodules, hilar and mediastinal disease, cardiac failure, emphysema, or pneumothorax), and 50 normal cases. The digitized images were displayed on a 1024-line monitor, and an ROC study compared performance of 4 observers with original film. They noted a slightly improved performance of digital images in detection of

hilar disease, but a significant reduction in performance with the digital images in all other cases; for example, the diagnosis of pneumothorax was made correctly with 86 per cent of the conventional radiographs, but only 61 per cent of the displayed images.

Seeley et al (1987) studied 8 paediatric chest radiographs, of which half were abnormal (interstitial shadowing). These were digitized at 100μm (12 bits) and printed out at resolutions of 5.0, 2.5, 1.25, 0.625 lp/mm (equivalent to matrix sizes of 4K, 2K, 1K, and 512 x 512 over a 35cm x 43cm area). Fifteen observers assessed what they termed the "seeability" of normal/abnormal structures on the hard copy printouts. They concluded that a resolution of 2.5 lp/mm (2K) was necessary, and might supply adequate resolution for a total digital radiology department. They were optimistic that this was achievable with existing technology, and did not comment on any difference in performance that might be expected between digital hard copy and digital displays.

Looking away from the chest, Kastan et al (1987) studied 10 films taken from air contrast examinations of the large bowel, of which 5 showed mucosal changes of inflammatory bowel disease, and 5 were normal (panel consensus). The images were digitized and printed to hard copy with pixel sizes 0.1, 0.2, 0.4, and 0.8mm. The images were viewed by 10 observers. Highest diagnostic performance was obtained with the highest resolution images, but the

improvement beyond 0.4mm was not statistically significant. They concluded that this was within range for resolution on available digital display monitors, but that resolution requirements were probably less important in inflammatory bowel disease than for the chest.

Dwyer et al (1987) conducted an ROC study of 50 chest radiographs from biopsy proven cases of interstitial lung disease, and 50 normal radiographs, digitized at 4000 x 4000 at 12 bits per pixel, printed out as hard copy at differing resolution. They found little or no additional benefit in the 4K matrix compared with 2K, but concluded that 2K was probably necessary. Templeton (1988) suggested that images provided by 1024 x 1024 x 8 bit monitors were still "very good", whatever that means.

Goodman et al (1988) conducted a further study of 35 chest radiographs, mostly selected from their previous investigation. They included 5 of their most frequently misdiagnosed cases of pneumothorax, and three cases of each of the following: hilar disease, pulmonary nodules, interstitial shadowing, pulmonary oedema, and emphysema; there were fifteen normals. The radiographs were digitized at 1680 x 2000 (0.2mm pixel, 2.5 lp/mm), printed onto hard copy at pixel sizes of 0.2mm, 0.4 mm, and 0.4mm with edge enhancement. Eleven observers took part. Performance improved significantly with smaller pixel size. For their five pneumothorax cases, detection rates were 89% (0.2mm pixel), 85% (0.4mm pixel enhanced), and 65% (0.4mm pixel);

for these images on the monitor display (1024 lines, 0.4mm pixel) detection rate had been 35%, but the observers were different. Strangely, the performance of this second group of observers with the original radiographs was not reported (for the first group, the detection rate had been 80%). The substantial difference in performance between displayed and printed images of nominally the same resolution was remarked upon.

A study by MacMahon (1988) used 60 chest radiographs, digitized to a pixel size of 0.1mm at 10 bits per pixel, to compare performance between displayed and hard copy digital images at the same matrix size. Images were displayed on a 1023-line 30 Hz interlaced monitor (pixel size 0.25mm) or printed onto hard copy (pixel size 0.2mm). Twelve observers took part, and they were not permitted to alter window settings on the display. Diagnosis was determined by a consensus of two radiologists, and each hemithorax was scored separately for normality or for one or more of the following: pulmonary nodule, pneumothorax, interstitial shadowing, bone lesion. ROC curves were generated for each type of lesion. There was little difference between the two formats for pulmonary nodules, but for the other lesions the hardcopy printouts of images scored significantly higher. They commented that the assumption that such images were equivalent for diagnostic purposes had hitherto been widespread. Reasons cited for the difference included the superiority of hard copy with respect to spatial resolution, and brightness (there are

also others). It is important to bear in mind that laser film recorders are capable of printing 4K x 5K for a CXR sized film (Cook et al 1989). These results point to the need for careful evaluation of display systems themselves, since matrix size is not the only parameter affecting the quality of the image.

Sheline et al studied forty chest radiographs, of which 15 were normal and 25 were from patients in whom a pulmonary nodule had been missed initially, but later confirmed by biopsy The images were digitized at 200μm and displayed on a 1000-line monitor. Six observers took part. Results showed improved nodule detection for the displayed image. They commented that the MacMahon study (1988) had failed to allow observers to alter window settings, thereby preventing exploitation of the superior contrast resolution of such systems.

Murphey et al (1990) conducted an ROC study of fifty-six radiographs of nondisplaced or minimally displaced fractures of the extremities and an equal number of normal images. These were digitized to produce spatial resolution varying from 5.75 to 0.72 lp/mm, corresponding to pixel sizes ranging from 0.087 to 0.694 mm. The conventional and digitized images were examined by 10 radiologists. They found a progressive improvement in observer performance as pixel size decreased. A pixel size greater than 0.16 mm (3.125 lp/mm) resulted in a significant loss of diagnostic accuracy in comparison with conventional radiographs.

Specific fractures in which a larger pixel size adversely affected the evaluation included torus injuries, corner fractures in child abuse, minimal avulsion injuries, and fractures that demonstrated only trabecular disruption.

Hayrapetian et al (1989a and 1989b) were the first group to publish results of work with 2048-line TV display systems, though they have so far reported on only 31 clinical cases.

Hayrapetian et al (1989a) digitized 31 chest radiographs to a matrix of 2048 x 2048. There were 6 normal images, and the remainder showed nodules, septal lines, or both (panel consensus). The images were shown to four observers on a 2048-line display system, as 2048 matrix hard copy print out, and as original radiographs. No significant differences were observed between the formats. It was concluded that results produced by 2K matrices displayed on 2K monitors were very close to results achieved with film and the traditional X-ray viewing box. They stated that further studies using larger numbers and more subtle images were now necessary. They also studied 25 "normal" chest radiographs from an anthropomorphic phantom, and 25 showing single simulated pulmonary nodules (1989b). The images were digitized to a matrix of 2048 x 2048 at 8 bits per pixel, and displayed on 512-, 1024-, and 2048-line monitors. No statistically significant difference was observed between the 2048-line displayed images and original film. Working with the same group, Hansell (1990) reported a study in which diagnostic accuracy for simulated

lung nodules tended to be slightly better with the display than for hardcopy, again using 2K images displayed on 2K monitors.

Cox et al (1989, 1990) described a further study with a 2000-line display system, in which 163 images were examined by six observers. The images were digitized at 4000 x 4000, and printed out at 2048 x 2048, or displayed at 2560 x 2048 x 12 bits. 64 cases were normal, and the remaining 99 demonstrated one or more of a variety of radiological features. ROC curves were generated for each pathological entity and display format. They concluded that the digital display formats were equivalent or better than conventional film in detection of costophrenic angle blunting, hilar and mediastinal masses, atelectasis, consolidation, parenchymal masses, and apical scarring; the improved detection of parenchymal masses was statistically significant; but performance was significantly worse for detection of pneumothorax and interstitial disease. Digital hard copy performed significantly better than the displayed images. They commented that it would be premature to conclude that "2K" display systems are equivalent to film for all detection tasks.

In the largest series to be published, Slasky, Gur, et el (1990) from the University of Pittsburgh studied a series of 300 chest images, using a team of seven observers. The images were digitized at 2048 x 2400, and printed out on film at full resolution or displayed at a matrix size of

1536 x 2048 at 8 bits per pixel. They included interstitial lung disease, pulmonary nodules, and pneumothorax. They found a significant reduction in performance for the detection of interstitial disease and pneumothorax when any of the digital images were used rather than the conventional radiographs. Although no radiologist in the study performed better on any digital image than on film, differences for the images showing nodules was not significant. The latter observation provides further confirmation that high spatial resolution is less critical for detection of pulmonary nodules, as other authors have similarly noted or predicted (Seeley et al 1978, Foley et al 1981, Goodman et al 1986, Sheline et al 1989, Newell et al 1988).

Slasky et al noted that poor performance was most marked with abnormalities requiring detection of information with a high spatial frequency component (e.g. fine lines); a pixel size of 0.2mm may not be adequate for primary diagnosis of these abnormalities. Kundel (1986) discussed the relationship between increasing spatial resolution in an image and the other features inherent in a digital system. He suggested that an image receptor with a greater dynamic range or better signal to noise ratio than a conventional film-screen system might yet produce images of good diagnostic quality, despite less spatial resolution; for example the contrast between air in the pleural space and the lung might be enhanced, making the identification of the sharp pleural line unnecessary for the confident

diagnosis of a pneumothorax. Slasky et al concluded, however, that the reduction in spatial resolution is not adequately compensated for by improved contrast resolution in such tasks.

Slasky et al further noted that although the results in their study were based on digitized film images, their findings in respect of the pneumothorax images agree broadly with a study using image acquisition with computed radiography (Fajardo, 1989).

In the Fajardo study, eight radiologists were shown film-screen chest radiographs and corresponding computed radiographic hard copy for a prospectively obtained series of 25 patients with pneumothoraces and 25 with other (or no) abnormalities. Observers scored images for the presence or absence of a pre-defined list of 11 abnormalities; they were deliberately not informed that the purpose of the study was to evaluate detection of pneumothoraces. Pixel size for the CR system was 0.2mm, and edge-enhanced CR images were also shown to the observers. The CR images performed significantly worse than the conventional radiographs. Intriguingly, the difference in performance was not related to the size of the pneumothorax.

Slasky et al commented that CR images have different characteristics from film radiographs that are not necessarily better; and similar experiments with other

pathological entities now need to be performed with CR. If the results are similar, they state, it will be necessary "to re-evaluate many PACS-related issues".

2.3. USE OF ZOOM

Most display systems are capable of some kind of zoom operation, whereby part of an image can be magnified and displayed at a higher resolution than can be used for the whole image. It is generally possible for the user to scroll through the magnified image, though the speed and ease of navigation through this image may vary considerably between different systems.

In a number of the reported series, 1K monitors were used to give a 1K "window" into a 2K image. The same approach may be applied to 2K monitors, providing a 2K "window" into a 4K image. Studies that have allowed the use of zoom include Kundel et al (1987) and Sheline et al (1989).

Some authors have assumed that, for diagnostic purposes, such images enable the same diagnostic interpretation to be achieved as a higher resolution display capable of showing the whole image in a single frame at full resolution. Certainly, manufacturers encourage this view, and this has obvious implications with respect to the costs of such systems.

The question of whether this is adequate, however, is unresolved, and it is possible that 1K monitors may not be suitable for viewing 2K images. The important issues in this regard concern the following:

- The role in the diagnostic process of landmark information within the image, orientating the observer and directing attention to prospective locations of abnormality, and the perceptibility of such information in the zoomed image.

- The fact that structures important to the diagnosis may extend over several magnified fields.

- The effect of limitations imposed by the user interface, in restricting easy scrolling around the entire image.

- The need to ensure that all relevant portions of the entire image are actually viewed at high resolution.

- Little is known about the psychophysical consequences of viewing an image in segments, but published work appears to indicate that performance is impaired: Carmody (1980) showed that scanning segments of images piecemeal resulted in an increase in false positives and reduced overall lesion detectability; he concluded that the inability to make rapid comparative fixations across different parts of the image interfered with the ability to reach a conclusion that the image was normal.

■ Radiologists tend to reach rapid judgements about images; or at least appear to be able to cull much of the information they need or want from an image on a brief inspection. Gale et al (1990) observed that 200ms glimpses of radiological images were generally sufficient for diagnosis - with not much difference in lesion detection rate when the time available for viewing the images was increased: they concluded that previous studies have underestimated amount of information obtainable from a single fixation. Since the full-sized, low resolution image is the first image that is usually presented to observers, it is possible that diagnostic judgements may be biased by the fact that such images will tend to appear normal; the zoomed image may only be inspected after such a judgement has already been made, or may even be ignored (as may have occurred in our own experiments - see page 201).

■ The consequences for perception of removing access to the complete image at full resolution are unknown, and there are many radiologists for whom the prospect of "radiology through a keyhole" is not an attractive one.

2.4. PSYCHOPHYSICAL FACTORS

Closely related to issues like these is an area that will undoubtedly require much more attention in the future: the psychophysical aspects of viewing displayed images.

Psychophysics is the study of the relationship between the physical attributes of the stimulus and the psychological response of the observer (Kundel 1979, 1986). Kundel examined the interplay between aesthetic appeal and diagnostic content, and pointed out that although spatial and contrast resolution are of utmost importance, aesthetics play a large part in the acceptance of an image. The acceptability of a display system may ultimately hinge as much on aesthetics as on any measurable parameter.

By habit and training, today's radiologists are well accustomed to film as the medium of primary diagnosis; the collective experience of diagnosis from display systems is small. In many centres, digitally produced images such as CT are printed onto film for the final diagnosis, simply from the preference by the radiologists for the film format - and convenience of image handling may be the most important reason.

In the Fajardo study referred to above, it was noted that the four observers who performed best in the task of detecting subtle pneumothoraces on film radiographs, performed worst with the CR images. The authors noted that spatial resolution could clearly not be the only factor at play. A possible explanation might be the presence of other, subtle diagnostic clues that are more easily represented on film, and that are more readily exploited by the most proficient radiologists.

Arenson et al (1990) point out that there is an important difference between the actual spatial resolution of an image and the perceived image sharpness; and similarly, between the contrast rendition and perceived image contrast; many of the differences can be accounted for by an understanding of the behaviour of the human visual system. For example, perceived image sharpness depends on the spatial frequency response of the visual system, which reaches a peak at around 5 cycles per degree, decreasing at higher and lower spatial frequencies; this non-linear behaviour explains why it is sometimes difficult to notice large lesions on an image without increasing the viewing distance - a phenomenon familiar to all radiologists.

Kosslyn (1989) stated that data displays will be effective only when they respect the properties and limitations of human perceptual and memory abilities. Carboni et al (1989) measured the effect on performance when observers were compelled to use a single display monitor to compare sequential images; they found a significant reduction in performance when there were delays between images even as short as 0.25 seconds, and further impairment took place if the screen went blank between images. The observations of Gale et (above) are also relevant in this context. It is likely to be many years before electronic display systems will be able to deliver the instantaneous performance that will make the most of observers' abilities.

There are many physical factors that cause difficulty when diagnosis is attempted from displayed images, and reference has already been made to observations regarding the superiority of digital hard copy over displayed images at the same resolution; reasons for better performance include: higher absolute luminance, greater perceived dynamic range, and better spatial resolution. The display luminance of monitors is an order of magnitude less than that of the viewing box. Extraneous light therefore has a much greater effect (Kundel 1986). It is not possible to increase the light intensity of the TV system beyond a certain level, because this reduces resolution due to "blooming" of the scanning electron beam. Kundel also observes that there are anomalies of the human visual system that must be taken into account, such as reduced retinal sensitivity to low contrast under conditions of low luminance.

2.5. MONITORS

Current state-of-the-art in monitor design is a new generation of 72 Hz 2560 x 2048 19-inch non-interlaced display systems (one manufacturer has set itself an objective of 80 Hz with a 23-inch screen, for the near future); they have been developed by small, specialist manufacturers, but are not yet supported as part of an integrated network by the current major PACS vendors.

Clark & Kriz (1989) discuss the problems of displaying images on such monitors at 12 bits per pixel. Although the human eye cannot differentiate the 4096 shades of grey from a 12 bit image, radiologists can certainly tell the difference between images from film digitized at 10 bits per pixel and film digitized at 12 bits - and subjectively prefer the latter. The human eye is approximately a logarithmic sensor, and wider dynamic range is more important than grey scale accuracy. Current display systems generally operate at 8 bits; it would seem that a display system with at least 12 bits at every stage from acquisition to display would be more appropriate. The current solution is to map 12 bit images through a lookup table to 8 bits; if the mapping is RAM-based, it can be changed easily by the user - window width and level manipulation. Development of the first 4096 x 3300 monitors is almost complete.

At 12 bits, a 2.5 x 2K display has 7.5 Mbytes data per image. Most existing systems take several seconds to load. Large capacity buffers and rapid processing are needed.

Film images viewed against a bright viewbox have a typical luminance of 800 ft-L (foot-Lamberts[1]), bright enough for the eye to be able to able to discount any inhomogeneity in the image. For current monitors, most fall in the range 30-80 ft-L, and 2K monitors with a luminance of 50 ft-L are

[1] Figures are quoted here using the units employed by manufacturers, rather that the preferable SI units.

now available. Monitors display fewer levels of gray, in a manner that may not be linear; resolution in horizontal and vertical axis may be different.

One of the limiting factors for displayed image resolution is veiling glare (scatter in the glass face plate of the monitor) which reduces modulation transfer (Blume 1990). Spatial noise (phosphor granularity) is significantly larger than temporal noise, and spatial noise limits low-contrast detectability.

There is a further problem for users of display systems, and one that is freely admitted by manufacturers (Blume 1990): manufacturers provide little quantitative information about performance. No test patterns have even been standardized, and the SMPTE (Society of Motion Picture and Television Engineers') test pattern is not really adequate, though an initiative by a European collaborative project (ISCAMI) is currently addressing such problems.

As monitor technology improves and differences in performance become smaller, larger studies will become necessary to demonstrate them.

While using monitors to display static images has problems, it should not be forgotten that there are many imaginative uses that may be made from their ability to display dynamic images: cardiac motion, flow, three-dimensional reconstructions, multi-modality image registration, etc.

2.6. FILM

Despite worldwide interest in digital technology, film sales are firmly on the increase; annual growth rate in the USA is 1% US, and as high as 5% elsewhere; worldwide, annual sales growth for video and laser hardcopy devices is 25% (news report, Diagnostic Imaging Inernational, April 1990).

McMillan et al (1989) have pointed out that it sometimes seems surprising that so many people are moving towards implementation of digital systems. New radiographic films have silver halide grains with greater surface to volume ratios; new intensifying screens offer improved speed: new film-screen systems have more latitude, improved contrast, better resolution at lower dose, and reduced noise. Spatial resolution for film lies between 5 - 8 lp/mm. Double emulsion radiographic film has a greater dynamic range than the single-sided films used in video or laser hard-copy output devices. Furthermore, silver prices are currently at their lowest level for years.

It should be recognized that, in the contest with film, digital technology is not the only area in which advances are taking place.

2.7. IMAGE PROCESSING

The advantages of image manipulation and processing have found application in radiology mainly in the context of CT and Digital Subtraction Angiography (DSA), but in other areas benefits have not yet been clearly identified.

Ishida et al (1984) conducted detection experiments using simulated low-contrast radiographic patterns in conjunction with a high-quality digital image processing system. The original screen-film images were processed to enhance contrast. The detectability of simulated patterns demonstrated a significant increase compared with the original images.

Oestmann (1989a, 1989b, 1989c) showed that various post-processing algorithms tested on digitally acquired images of simulated nodules (on volunteers) did not improve diagnostic accuracy; edge enhancement improved the detection of simulated lines on CR images to a point where they could be seen as well as with conventional film; and that dual energy subtraction (with CR) could be used to compensate for reduced detection of simulated pulmonary nodules.

In an earlier study Oestmann (1988a) also showed that post-processing of digitized (1K) film images significantly reduced the diagnostic accuracy of subtle lung cancers (though in this study the images had been photographed from

a display, and were presented to observers by projection onto a slide viewer!).

Barnes et al (1989) used simulated pulmonary interstitial patterns shadowing in experiments with bone subtraction from chest radiographs; results from cancelling bones in 1K digital hardcopy images showed no benefit.

Burgess (1988) considers that "there is certainly no convincing evidence that image processing techniques will be beneficial for enhancing display of digital medical images. The one possible exception to this is contrast transformation techniques."

2.8. DATA COMPRESSION

High resolution digital images constitute large volumes of data (Table 1, page 42; Table 2, page 43). The amount of data - determined by the matrix size and the number of grey levels within the image - directly affects transmission times and storage requirements. The use of data compression algorithms reduces the volume of data.

Despite the widely acknowledged practical importance of data compression, particularly as higher resolution images are contemplated, it is interesting to note that none of the observer performance studies quoted above have included assessment of image quality considerations from the use of compression algorithms. In fact, very few studies have

been published, and these have been largely limited to conference papers and practical demonstrations of the kind of results that can be achieved.

This is not surprising. It is difficult to consider applying methods that might conceivably impair image quality at a time when the performance of digital images still lags significantly behind the performance of conventional film. Definitive studies on data compression need definitive algorithms to study, and these in turn need definitive digital images and display systems.

The level to which data can be compressed depends on many factors, such as image type and quality, matrix size, and compression algorithm. Images with large homogenous areas are most suited to compression; Huang (1987) states that, for typical radiological images, the larger the image matrix, the more suitable it is for compression.

Results vary from centre to centre. A study carried out by Lo (1986) concluded that visual degradation was not excessive when using a compression ratio of 16:1.

Halpern et al (1989) reported a study based on applying a simple quadtree-based compression to digitized radiographs from 100 urograms. Each image was digitized on a "1K" matrix and displayed to 4 observers on a "1K" monitor at nine decreasing compression ratios ranging from 90:1 to 4.2:1. Sensitivity decreased with increasing compression

ratios at and above the 11:1 level. No loss of sensitivity was noted with a compression ratio of 4.2:1. Sensitivity decreased more precipitously for calcification than for soft-tissue masses. The authors concluded that quadtree compression ratios above 4.2:1 may result in loss of sensitivity to clinically relevant findings.

There are two main radiological research groups with the resources and technical expertise to develop, test and implement their own compression hardware and software, without reliance on manufacturers: Huang and colleagues at UCLA, and Dwyer and colleagues in Kansas. This inevitably introduces a bias of emphasis that reflects the approaches they have followed.

Ho et al, of the UCLA group (1987) expressed the view that image compression at a ratio of approximately 10:1 would be essential for economically feasible PACS to be implemented. The ethernet link between their viewing stations had reached transmission speeds of 1 megabit per second. This meant that it took eight seconds to transmit a "1K" image, or 32 seconds for a "2K" image, resulting in long delays retrieving images directly from the central archive, with restriction of the number of active viewing nodes due to heavy competition for network devices. They explored image compression at a ratio of 10:1, and reported work with a full frame fast cosine transform technique (FCT), using hardware of their own construction. By using concurrency of the FCT process and direct memory access design for data

transfer, the compression module could compress or decompress a "1K" image in one second and a "2K" image in four seconds. Block quantization techniques leave blocky artifacts that many consider unacceptable for diagnostic purposes. FCT does not suffer this degradation. Image quality studies using ROC analysis had demonstrated integrity of diagnostic information with this algorithm at compression ratios of 10:1 (though further details of the ROC results were not presented). Such an algorithm would make it possible for display stations to receive compressed images over a network and produce a viewable image after a tolerable delay. They presented a demonstration of the functioning compression module.

Although blocked compression algorithms might not seem an obvious choice for clinical use on account of the artifacts in the decoded image, ACR-NEMA in fact publishes a blocked DCT compression standard (ACR-NEMA 1990). Ishida et al (1990) have attempted to reduce the prominence of the blocky appearance of the image by applying a modified DCT algorithm to CR images of the chest. 30 images were displayed on 1024 x 1536 video display systems at compression ratios of 5:1, 10:1, 15:1, 20:1, and 25:1, and ROC experiments were conducted using 6 observers: they concluded that 10:1 represented the upper limit of acceptable compression using their method.

Bramble et al (1987), of the Kansas group, identified three main types of data compression relevant in the context of

diagnostic radiology: data-preserving or lossless compression; data-losing or "lossey" compression (both of which are algorithmic); and "clinical" (which consists of discarding less useful images, image portions, etc.). They also summarized three main parameters by which compression algorithms should be judged:

1. Compression ratio - expressed as original:compressed

2. Computation time - increases with complexity of algorithm and compression ratio.

3. Image quality - which becomes a factor only in data-losing algorithms (in lossless compression, the decoded image is by definition mathematically identical to the original): images quality must then be evaluated by ROC analysis.

In an investigation of lossless compression, they studied over 1000 images using Huffman and Differential Shift encoding. Compression ratios of 1.5 to 3:1 were achieved, but encoding and retrieval times ranged from 221-765 seconds: computation time was found to be directly proportional to the square of the digital array size.

Background elimination was explored with CT and MR images; 5:1 compression ratios were easily achieved, with no implications for image quality. "Clinical" compression was also explored in various settings; radiologists are

reluctant to discard images, but it was felt possible to achieve reductions in data storage of 4:1 to 5:1 - this however requires more effort than most radiologists are probably willing to make.

Ho et al (1987) used an FFBA algorithm (full frame bit allocation), with a Konica laser film digitizer and a 3M laser printer. 31 chest films were studied, of which 18 showed fine septal lines, 14 showed fine nodularity, and 6 showed neither; films were digitized to "2K", and compressed 11:1 (40 secs to encode or decode). Each observer saw the original film, the digitized film, and the digitized compressed film, at 10 day intervals.

Their conclusion was that slight image degradation occurred, affecting septal lines less than nodules; but that the loss was not significant on ROC analysis.

MacMahon et al (1991) conducted observer performance studies using a proprietary adaptive blocked cosine transform algorithm applied to chest radiographs. They used 60 radiographs showing pneumothorax, interstitial shadowing, nodules, and bone lesions. The radiographs were digitized at a matrix of 2048 x 2048, to a bit depth of 10. The images were printed onto hard copy at the same resolution, without compression and at ratios of 25:1 and 50:1, and the digital images were presented to 12 observers. They demonstrated a reduction in performance relative to the non-compressed images for detection of

pneumothorax and pulmonary nodules, but these differences failed to reach significance for the 25:1 compression ratios. The reduction in performance for the higher ratio was significant. For the other types of pathology, differences were observed between the compressed and non-compressed images, but these did not reach significance. They concluded that a compression ratio of 25:1 might be acceptable for primary diagnosis, and would almost certainly be acceptable for archiving purposes.

Hierarchical algorithms such as the S-transform are lossless, and allow image encoding and decoding to take place in progressive stages; thus the decoding time is mitigated by the appearance of a preview image on the screen, almost instantly; this is currently a promising area of research, especially since lossless compression is preferred by manufacturers for medico-legal reasons (see page 249), who ultimately may give their users little choice in the matter (H Blume, and T Wendler, Philips Medical Systems, personal communication).

As technology develops, transmission speeds will increase and storage devices will gain capacity, and the need for compression may recede.

2.9. INTELLIGENT WORK-STATIONS

Given current data transmission rates, and the likely high demand for image data content, an important requirement for

successful diagnostic workstations will be the ability to "pre-fetch" relevant images, such that a local buffer can be preloaded with all or most of the images likely to be required for a particular reporting session.

The concept of the "intelligent" work-station has accordingly evolved, and will be crucial to the successful integration of PACS workstations into clinical radiological practice. Among the earliest practical implementations of such systems is the work of Levin (1990), who described a rule-based pre-fetching expert system for dearchiving comparison images in nuclear medicine. Similar work has also been proposed and debated by other groups, such as Wendler and Wein (1990) and van Poppel et al (1990).

Wendler has also theorized upon the importance of the user interface, which has an significant influence on some of the psychophysical parameters discussed earlier. However much one might wish to modify the handling and functionality of a workstation, however, most users lack the resources to make changes themselves, and are therefore totally dependent upon what the manufacturer is prepared to provide. There has been a recent trend towards use of touch-screen monitors to control image handling functions: one advantage of such an approach is that upgrading such systems becomes less dependent on hardware; refinements to the interface can be performed by software modification.

2.10. CLINICAL MATERIAL

2.10.1. Chest images

It will be noted that the majority of observer performance studies have focused on chest imaging: the plain chest radiograph is by far the most frequently conducted examination in diagnostic radiology, accounting for 40-50% of the examinations performed in a typical radiology department, and it has high requirements for contrast and spatial resolution. Until successful results are obtained with chest imaging, digital implementation for general radiology will not be achievable.

Film is not an ideal receptor of the chest image because it does not have the latitude to record both lung and mediastinum without loss of image quality (Goodman 1988). Fraser (1989) has noted that primary X-ray transmission through the lung is approximately 50 times higher than through the mediastinum, but that the useful exposure latitude of film is only a factor of 5, making it extremely difficult to obtain good anatomical detail from both on a single image. The ability of film to record a usable image of the mediastinum depends largely on scatter from the lung fields. High kVp technique reduces visibility of skeletal detail and calcification.

Most series have used either lung nodules to evaluate low contrast resolution, or septal lines and pneumothoraces to

evaluate spatial resolution performance; it should be noted that the spatial resolution requirements for amorphous parenchymal pathology are still largely unexplored.

Digital technology is not the sole thrust of research efforts to improve chest radiography: the recent development of the AMBER system (advanced multiple beam equalization radiography), which uses a scanning slit to achieve exposure equalization across a chest radiograph, is now undergoing evaluation (Vlasbloem 1988). Its use in conjunction with dual energy CR is also being explored, and future developments include the use of kVp equalization techniques (Shultze Kool 1990).

2.10.2. Breast images

A small number of studies have also been carried out on breast images. The detection of microcalcification depends on high spatial and contrast resolution. A study carried out by Kimme-Smith (1989a) compared conventional and (magnified) microfocal mammograms with video-digitized enhanced images (acquired using a 512-line vidicon system). The results showed that three radiologists experienced in mammography scored equally for conventional radiographs and digitized images and significantly higher on macro-mammograms. Three inexperienced radiologists scored very poorly with the digitized images. She concluded that such images should not be used by those without extensive mammography experience. However, this series included only

10 examples of malignant microcalcification, compared with 21 examples of benign (much coarser) calcification, and lacked statistical power.

She compared the video digitization/enhancement technique with CR-acquired and processed images on hard copy (Kimme-Smith 1989b) and concluded that this provided better image quality than CR without enhancement, but that both methods had serious deficiencies in comparison with film.

An earlier study by Smathers (1986) compared xerograms, conventional radiographs, and radiographs digitized at 2K and displayed on 512 x 512 monitors (with zoom). He reported that xerography yielded the best performance, followed by the digitized displayed images, followed by the film-screen combination. The study was based on phantoms using specks of aluminium oxide and pulverized bone fragments of known size and number to simulate calcification. The bone speck size that corresponded to a 50 per cent detectability rate was 0.55mm for xerography, 0.573mm for the digital method, and 0.661mm for film. Clinical images contain microcalcifications as small as 0.2- 0.3mm, however, so there is clearly room for more investigation based on clinical material.

Chan et al (1987) digitized conventional film screen mammograms to 0.1mm; the digital images were printed out as hard copy at full resolution, and there was lower detectability of microcalcification in comparison with

original film. Oestmann (1988b) conducted a study using CR (hard copy) at 0.1mm resolution, using a breast phantom model with superimposed calcifications ranging in size from 50-800μm; 30 "normal" and 30 "abnormal" images were shown to 4 radiologists; the form and number of microcalcifications could be observed more accurately with conventional film, but no significant difference was found between the two modalities in terms of detection of cases where microcalcification was present.

2.10.3. Other pathology

In a study of images from patients with subperiosteal resorption, Murphey (1989) obtained magnification radiographs from 40 normal and 40 abnormal subjects. The images were digitized to 4K at 12 bits per pixel, with a 80μm scanning spot size, yielding a spatial resolution of 11.4 lp/mm (0.044mm pixels). A CLAHE contrast compression algorithm was applied, and pixel averaging was used to generate lower resolution images (down to 512 x 512); the digital images were printed out as hard copy, and were viewed by six radiologists. Murphey found a significant difference in performance of between the digital images at 2K (5.7 lp/mm, pixel size 0.088mm) or less, compared with film; the 4K images approached the resolution of film. He also observed that, in the case of radiography of the extremities, macroradiography would be a possible solution to the requirement for smaller pixel sizes with CR.

Fajardo (1987) reported a study in which CR had been used to obtain images from 100 patients undergoing excretory urography, in whom matching film radiographs were also taken. Pixel size was in the range 0.2-0.3mm for the CR images, which were printed out as hard copy and interpreted by 3 radiologists. He found no significant difference between the diagnostic conclusions drawn from the two systems.

Gross et al (1990) reported a comparison between original film and digital hard copy from 50 abdominal radiographs of neonates with necrotizing enterocolitis and 50 normal controls. 20 x 25 cm films were digitized to a matrix of 1600 x 2000 (i.e. a pixel size of 125μm) at 12 bits per pixel, printed onto film, and viewed by twelve observers. They reported no significant difference in diagnostic performance between film and the digital hard copy; however they chose not to use ROC methodology in processing their data, using instead a matched-pair comparison of a three-point non-parametric score.

Very few studies have been published using clinical material from elsewhere in the body, and the resolution requirements have not been documented.

[CR images are digitized at 100μm or 200μm, depending on size and type (standard or high resolution) of phosphor.]

2.11. ROC ANALYSIS

The need to use objective observer performance studies in this context is linked to current lack of knowledge about the way a radiological diagnosis is reached.

ROC analysis is based on statistical detection theory as described by Wald (1950). More recently, it has been applied to medical imaging systems (Swets 1979, Kundel 1979). ROC analysis requires the observer to indicate whether an image displays an abnormality (signal) or not. The observer must also indicate his certainty of his decision. A brief, simple explanation of the ROC curve is provided in Appendix 2.

In practice, formal ROC studies are time-consuming to perform. More importantly, they do not adequately simulate radiological reporting in a clinical situation. FROC (free response ROC) allows multiple abnormalities and observer responses per image, requiring correct localization of a lesion before a "true positive" response can be recorded (Chakraborty 1990). Although this comes closer to a realistic situation, the FROC approach, in common with LROC (location ROC), has so far only successfully been applied to multiple simulated lesions, in situations where there has been an equal probability of a lesion occurring at any location on the image.

ROC techniques tend to make good or decisive observers appear to perform slightly worse, and poor or indecisive may appear to perform better.

A further problem in selecting case material for ROC studies is the need to verify the diagnosis by some independent means: the validity of a test cannot be established by the test alone. Panel decisions are not always satisfactory, and often lead to exclusion of subtle or borderline material. Various authors have proposed methods and classifications for verifying the truth of diagnosis (e.g. Ker 1988). The prevalence of subtle lesions is small, so it is impractical to assess performance simply by monitoring clinical practice.

Studies still occasionally appear that are purely subjective in the assessment of image quality (e.g. Yonekawa et al, 1988).

2.12. ECONOMIC PROJECTIONS

The first feasibility study of a filmless department was published in 1983 (Huang, 1983). Huang stated that by elimination of film as an acquisition medium would reduce departmental expenditure by $400,000 per year, though he warned that film would still be needed for some purposes. This study however failed to take account of capital outlay. Although PACS would answer many current problems, the capital outlay would be enormous.

Beard et al (1990) modelled the performance of networks of different specification. They concluded that faster, higher specification networks are more cost-effective than lower speed ones, and might be expected to become cheaper than film-based systems by 1995. They foresee that PACS can become cost effective within the next 5-10 years. They demonstrated that workstation cost and number are the most significant factors in the overall cost of the network. Many workstations will be needed: the greater their sophistication, the greater will be the total cost.

There are difficulties in extrapolating cost analyses from the USA to the situation that currently exists in the UK. One important reason is hardware costs, which in the UK tend to be 75-100% higher for identical products. A quotation from one of the manufacturers in 1989 suggested that a hospital-wide system with 140 1K work-stations would cost £8,000,000.

Stockburger et al (1990) concluded that the direct economic benefits within the radiology department could not alone justify investment in PACS: increased productivity throughout a hospital, with the prospect of shorter patient stays would need to be taken into account - presuming that it actually happens.

Glass (1990) shares this view, and links economic viability to the total elimination of film expenditure and to the introduction of PACS in a single step, though he has also

stated that such cost analyses serve mainly to convince administrators to proceed with implementation, rather than to provide a realistic model of what will actually happen. Some of the assumptions in an economic justification for one attempt at total system implementation at a site in the UK include: ten-year lifespan for installed equipment with no provision for upgrading or replacement of hardware; acceptability of 175μm pixel size throughout, with display systems, processing power, and data storage provision calculated accordingly; annual maintenance costs held to 4 per cent of purchase value, over a ten-year period; the systems becomes economically viable in the tenth year of operation (Glass & Slark 1990).

In a computer simulation exercise, a software package has been developed and is commercially available (CAPACITY, BAZIS Ltd, Netherlands); it uses 500 variables (van Gennip, 1990) and simulation experiments tend to confirm that there will be a 7-10 year delay before a justification for PACS can be constructed purely in terms of cost savings.

2.13. DISCUSSION

The foregoing represents an attempt to identify and survey key issues that have been studied by other investigators in this field. It can be seen that image quality cannot sensibly be considered in isolation - there are too many other factors at play.

No clear picture emerges from the literature regarding the parameters needed for total digital radiology; some of the results are conflicting, and no real consensus has yet been reached regarding the necessary matrix size for images and monitors. The required parameters may vary for types of examinations and between pathologies.

Hansell (1990), amongst others, has pointed out that there has been decrease over the years in the estimated acceptable pixel size, from 1mm (Foley 1981), to 0.4mm (Lams & Cocklin, 1986), to 0.2mm (MacMahon 1986, Goodman 1986); Fajardo (1989) questioned if even 0.2 mm was adequate. Lack of agreement has been ascribed to factors such as use of gross rather than subtle lesions in observer performance studies, and the use hard copy rather than displays. It is also interesting to speculate whether some of these findings might have been influenced by the prevailing technological state-of-the-art when the studies were performed. It seems likely that the most subtle lesions need pixel sizes of 0.2mm or less.

It has been argued in the literature that most studies are based on images derived from film, and that their conclusions are not relevant; the digital system can only hope to do as well as the conventional system from which it is derived (Adam & Dawson, 1990; Arenson et al 1990). The Fajardo study (1989) was based on images acquired with computed radiography, however. Many observers still make the mistake of referring casually to computed radiography

as 'direct digital acquisition': it is not, and involves digitization, with a similar process of spatial and contrast quantization of optically derived data. CR has limitations, and use is not widespread. Prospective accumulation of suitable clinical test images, with adequate verification of the diagnosis, is even more complex a task with CR than it is for film. It is entirely reasonable that displayed image quality should be tested with images derived from film.

Kundel (1990) has observed that the highest resolution images can be of no practical value to patients unless the information from them can be communicated in a timely manner, and points out that this frequently does not happen with film. The timeliness of the digital link may arguably compensate for lack of resolution or reduced quality.

Other approaches are also being explored to help the observer extract the maximum information from the displayed image. Krupinski et al (1989) provided observers with visual feedback by showing them a map of their eye movements during inspection of images with subtle chest nodules. Performance improved.

Another approach is the use of image analysis, or computer aided detection. Numerous authors have described prototype systems for searching images for lesions such as pulmonary nodules or mammographic microcalcification; Doi et al (1990) provided a state-of-the-art practical demonstration

of systems for automated detection of such lesions as well as automated cardiac size estimation in chest radiographs and measurement of stenoses in digital subtraction angiograms; they have pointed out (Giger et al, 1990) that radiologists fail to diagnose pulmonary nodules in up to 30% of positive cases, and believe that such systems provide a valuable aid to the radiologist, directing attention to subtle lesions that might otherwise be missed.

It is clear that large studies using carefully selected clinical material and simulating the clinical environment, will be needed before we can really hope to understand the psychophysical consequences of widespread introduction of soft copy reporting. It is difficult to simulate by experiment the influence of the film-reader's art, and it would be a great loss to radiology if diagnosis from digital displays were to become a potentially sterile exercise in confirming or excluding some prior hypothesis for each image, with each image acquired under conditions designed purely to demonstrate the most likely lesion. Will display systems only be able to cope with certain radiological conditions rather than others? Will we be only able to find the conditions that we look for? The ability to make a diagnosis from an image that appears normal to the untrained eye is part of the radiolologist's stock in trade; it should not be forsaken lightly.

3. MATERIALS AND METHODS

3.1. EQUIPMENT

During the period of this investigation, two different display systems were subjected to examination. Both were installed at St Mary's Hospital and supplied and maintained by Philips Medical Systems (UK). The equipment was purchased with funds from the Department of Health.

Both sets of equipment were offered in the form of complete packages, with no option to vary their specification.

3.1.1. System 1

The first system was delivered in April 1988, having been promised for delivery in November 1987, and was finally handed over in July 1988 after many technical problems during installation.

The system is shown schematically in Figure 3.1, and consisted of:

- "1K" laser film digitizer (Mar-Scan) (Konica, Japan): (scanning spot approx. 100μm) with "1K" viewing console, and "1K" x "1K" memory (Mar-Trans).

Figure 3.1: System 1: schematic.

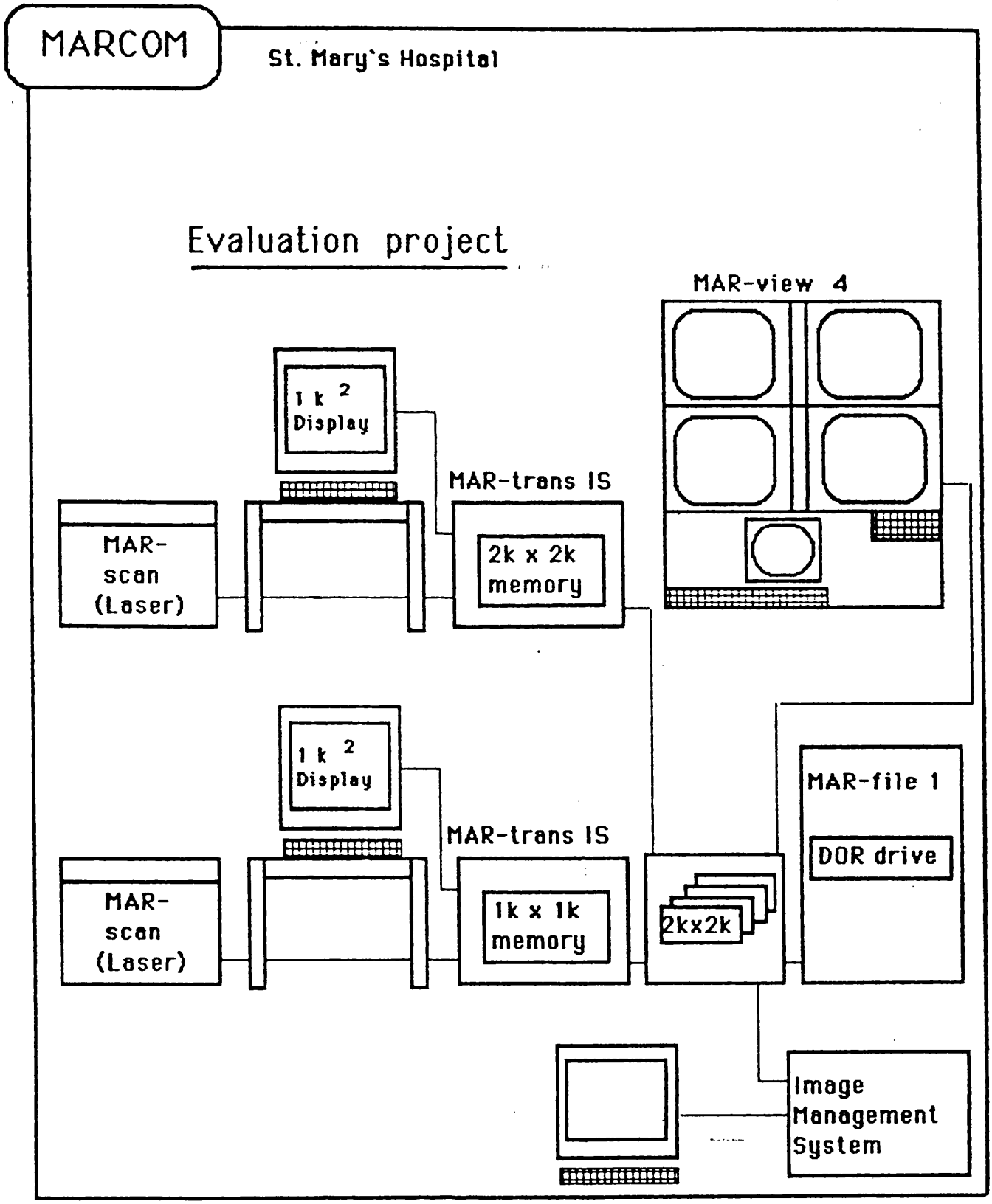
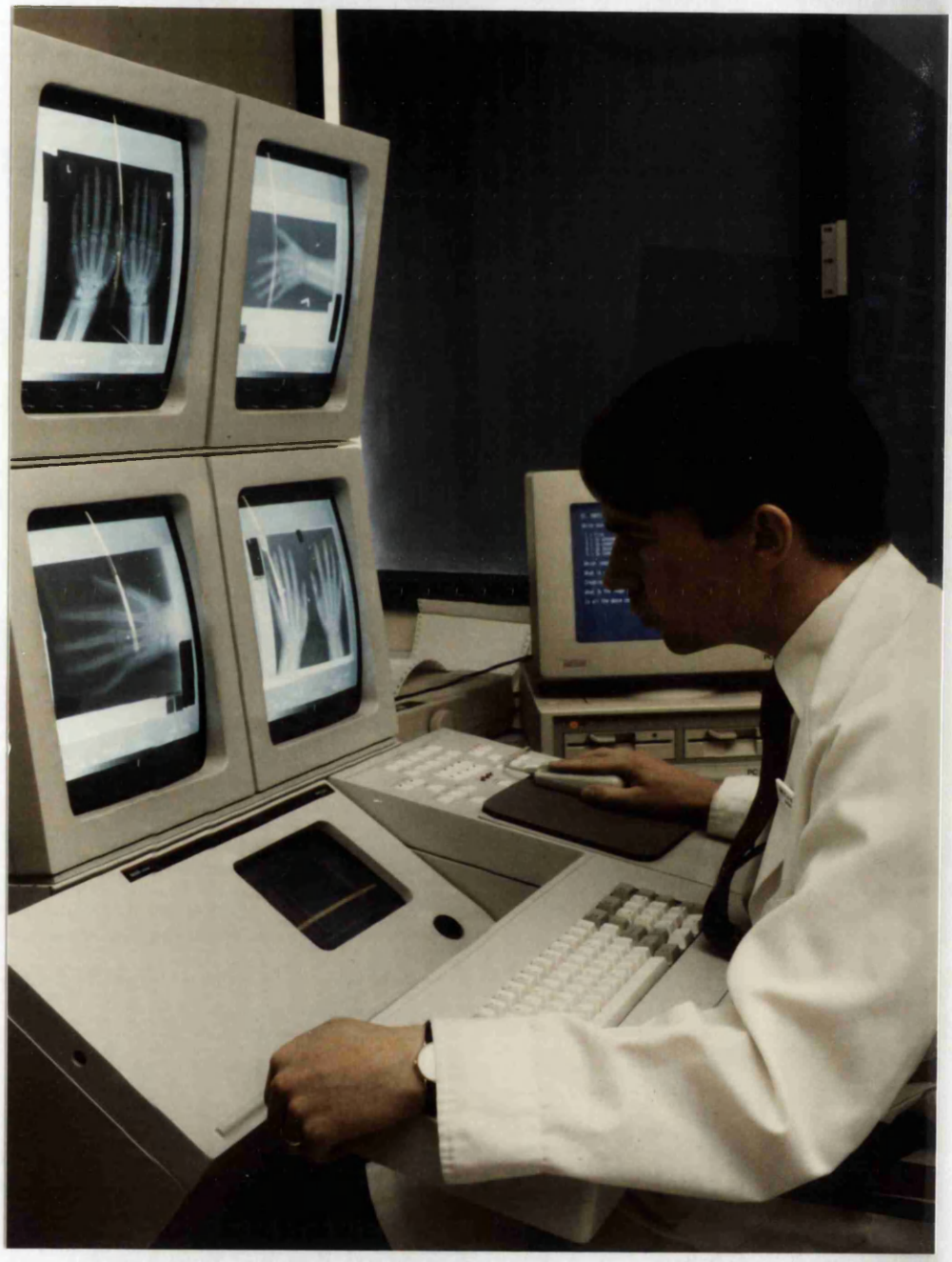


Figure 3.2: System 1: display system in use.



- "2K" laser film digitizer (Mar-Scan) (Konica, Japan): (scanning spot approx. 100μm) with "1K" viewing console, and "2K" x "2K" memory (Mar-Trans).

[The two scanners differed only in that they were configured to scan at different matrix sizes - nominally 1000 x 1000 ("1K") and 2000 x 2000 ("2K"). The bit depth for both scanners was stated by the manufacturer to be 8 bits per pixel.]

- Optical disk archive (Mar-File) with single 12 inch drive (1GByte per side)

- 160 MB local magnetic disk buffer; with archive terminal and printer

- Image management system (IMS) with terminal and printer (image management software/patient database)

- 3M digital laser film printer (added briefly)

- Display terminal: 4-monitor viewing station (Raytel Corporation, USA) (Mar-View) with local 160 MB magnetic disk [1024-line monitors, with band width corresponding to 48MHz pixel clock rate and 8-bit video coding, interlaced, vertical raster].

A proprietary blocked discrete cosine transform (DCT) algorithm was available on the system (approximate ratio, 15:1).

Part of the difficulty with the system was eventually attributed to the fact that it was originally designed to handle "1K" images, and the handling of "2K" image data sets was unreliable. This eventually necessitated its replacement.

The architecture and capabilities of the equipment imposed a number of practical constraints upon the design and structure of the observation experiments.

For example, the optical disk drive could only read a single side of a single disk at one time, and the disk had to be changed manually; "2K" uncompressed images require much archive space. In order to handle the VDU image sets for each group in a manner that would permit efficient digitization and rapid viewing, it was necessary for the images from each group to be archived on the same side of an optical disk. This made it essential to wait until each series of images was complete before digitizing and archiving any of the images from that series.

The magnetic disk drives of Mar-View and Mar-File had a capacity of 160 MB, equivalent to data from roughly 40 "2K" uncompressed images. This had significant consequences both for the digitization procedures and for viewing; a

group of 100 "2K" uncompressed images would have to be divided into three groups, each taking approximately 3 hours to transfer between the archive and the viewing console. Digitization, IMS, and archiving routines had to be devised to take account of these constraints, and these procedures meant that images from the "2K" uncompressed format would always be identifiable to observers as such.

Further time constraints were imposed by the fact that upgrading of the optical archive to the ACR-NEMA standard was anticipated in autumn 1988, after which data already archived would no longer be retrievable.

The display monitors had 1024 lines and could not display "2K" images at full resolution in a single frame; at "2K", only a quarter of the image was visible at once, though it was possible to scroll around the image while using this "zoom" facility, using the "1K" display as a window into the larger matrix image. It was possible, however, to display the entire image at reduced resolution.

Digitization matrix sizes are generally quoted for a full-sized (35cm x 43cm) film; the number of pixels per image at any given resolution therefore varies with film size: the actual matrix sizes that resulted are shown in Table 3 (page 136). All references to "1K" and "2K" are in fact approximations.

The addition of a laser printer to the system was for too short a period to enable a formal study of hard copy printouts in comparison with the displayed images, and in any case only a video interface to the hard copy device was available.

The following image manipulation facilities were available, and all observers were free to use them:

Magnification x 2
Magnification x 4
Window width
Window level
Rotation (+ or - 90 or 180 degrees)
Lateral inversion (to mirror image)
Grey scale inversion

The existing IMS had been designed to handle patient data; each "patient" was able to undergo any number of individual procedures, each of which might be associated with one or more images. There was thus a three-tiered hierarchy; efficient management of patient data takes place at the upper two levels, while efficient image management (fast retrieval and viewing, for example) takes place at the lowest. Evaluation projects need a rather different structure; the data to be managed relates to individual images rather than groups of images. The software of the system had to be modified accordingly.

As the cases were accumulated, they were entered into the IMS database and allocated file numbers for manual retrieval and random numbers to ensure random position within the group.

Figure 3.2 shows the system in use.

3.1.2. System 2

The components of the second system, which was commissioned in January 1989 and remains the current installation, are shown schematically in Figure 3.2, and are listed below:

- Two laser film digitizers (FD2000, Dupont, USA): (scanning spot size 210μm; bit depth programmable by user to 8, 10, or 12, and set to 12 throughout duration of our investigation).

- Optical disk archive with single 12 inch disk drive (1Gbyte per side)

- 180 Mbyte magnetic disk for local storage.

- 4 x 825 Mbyte magnetic disk buffer

Figure 3.3: System 2: schematic

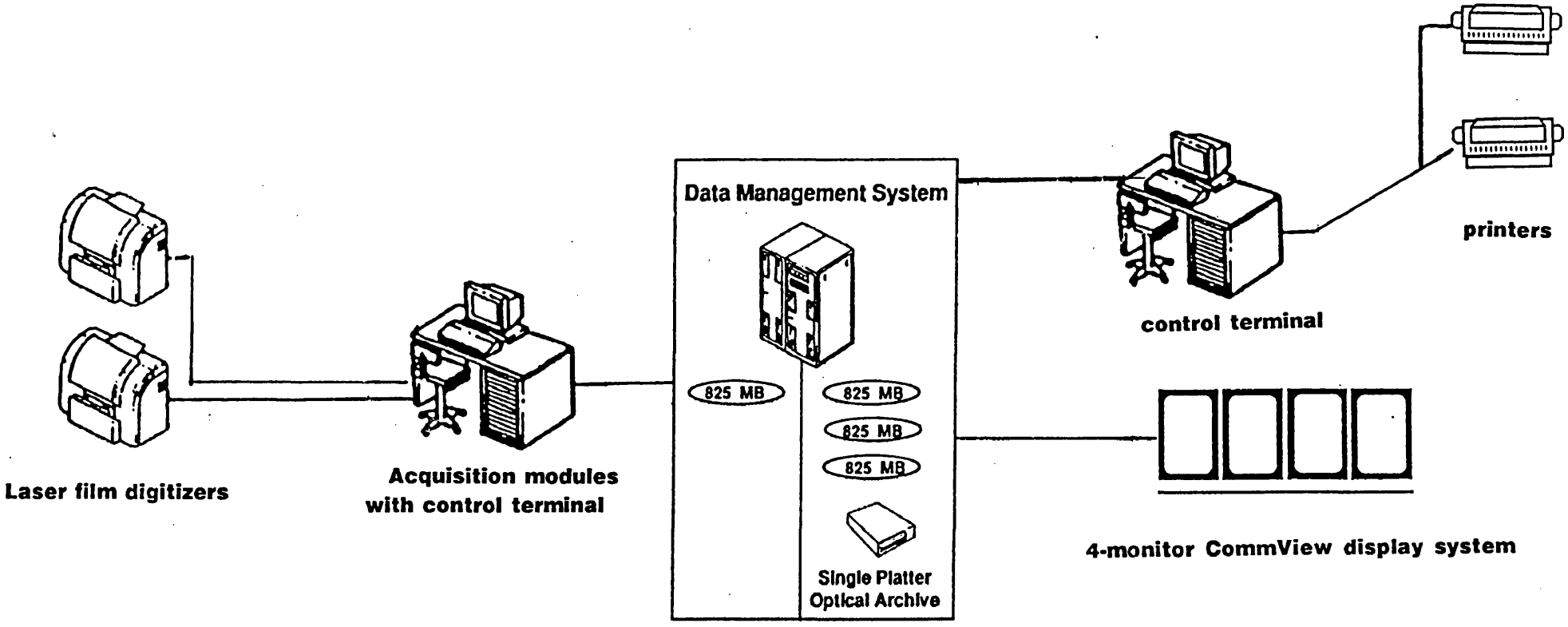


Figure 3.4: System 2: display system in use.

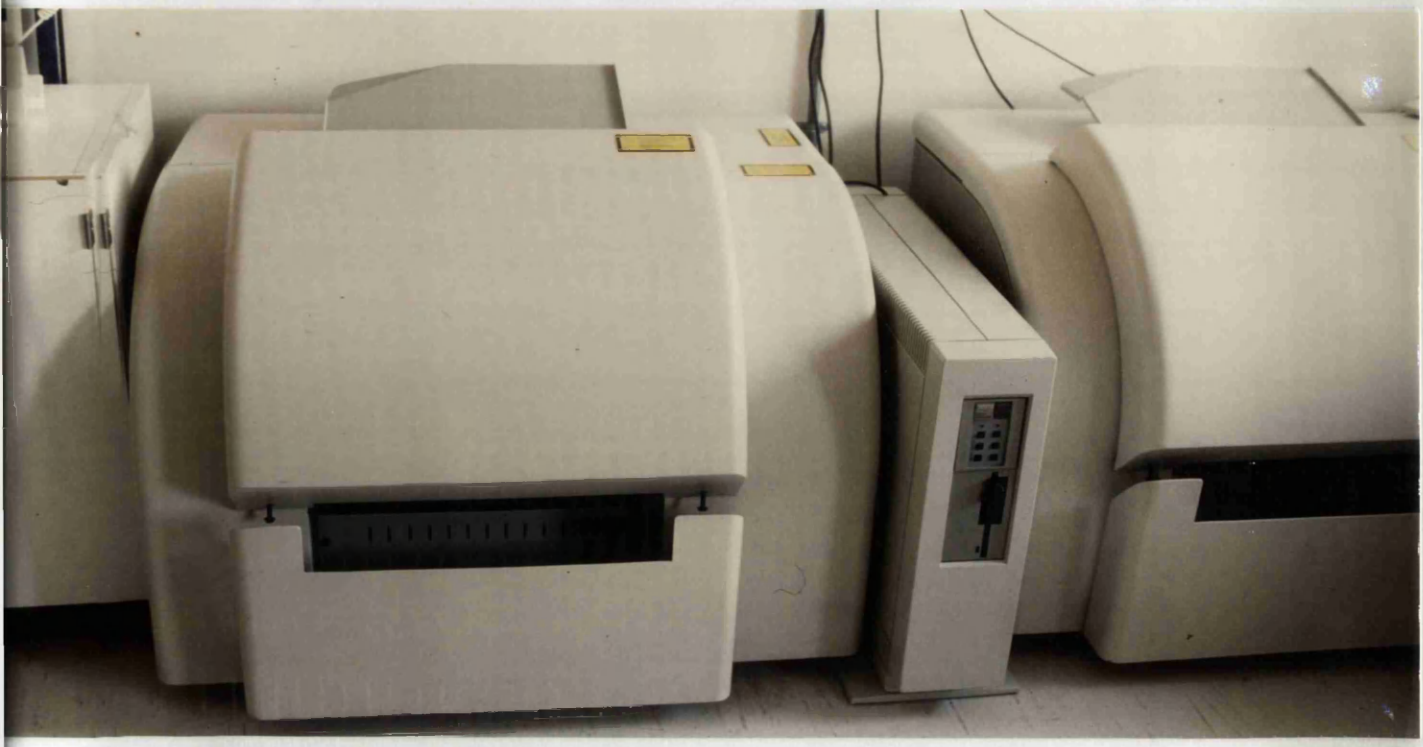


Figure 3.5: System 2: digitizers.

- Display workstation: 4-monitor display system, with separate text data entry terminal. [1280 line monitors with band width corresponding to a 60MHz bandwidth and 12-bit video coding, non-interlaced, horizontal raster]

- data communication link to manufacturer (via modem)

This equipment is sold under the brand name "CommView", in a joint venture between Philips Medical Systems and AT&T in the USA.

As with System 1, the combination of a "1K" display with "2K" digitization enabled "zooming", to view a portion of a "2K" image at full resolution.

The system also incorporates a proprietary irreversible compression algorithm comprising two steps, DPCM followed by Lempel Ziv encoding. The approximate ratio is part of the system configuration, and changes in ratio require re-programming in UNIX by service personnel or the system supervisor. Compression level "10" is equivalent to a ratio of 3:2, reversible compression. Compression level "1" is equivalent to a ratio of 7:1 or 8:1, depending on the particular properties of any given image. The intermediate levels represent intermediate ratios of irreversible compression.

In order to be permitted to use this algorithm, it was necessary to provide the manufacturer with a written undertaking that data compression would not be applied in any situation that might involve patient diagnosis or management, owing to the manufacturer's anxieties regarding medico-legal liability.

The approximate monetary value of System 2 was stated by its supplier to be £700,000 at 1990 prices.

The following image manipulation facilities were available, and all observers were free to use them:

Magnification x 2
Magnification x 4
Loupe function to magnify small areas of the image
Window width/level: using either a trackball or programmable preset settings
Rotation (+ or - 90 or 180 degrees)
Lateral inversion (to mirror image)
Grey scale inversion
Scroll facility, to navigate around the enlarged image.

3.1.3. Calibration and quality assurance

Great care was taken to ensure uniform performance of the four monitors, including uniformity of brightness and colour temperature (see page 183). The phosphor on one of

the screens proved to be from a different batch, and was noticeably different in colour; though within manufacturers tolerances, the supplier agreed to replace it.

Further quality assurance issues will be discussed elsewhere.

3.1.4. Additional equipment

An IBM-compatible microcomputer was used for data capture.

Further items of equipment for testing and quality control (such as a frequency spectral analyzer) are referred to under "Physical experiments" below.

3.2. CHOICE OF CLINICAL CASE MATERIAL

An important part of the evaluation of digital display systems for clinical use involves objective assessment using genuine clinical case material. In order to test a system to its limits, it is necessary to choose pathological conditions to study that are subtle, and difficult to diagnose even on film, and about which it is possible to create doubt in the mind of the observer.

A number of clinical conditions are suitable candidates in this context, some of which have already been studied, most

notably in comparisons between film and digitized images on hard copy rather than between film and displayed images. These are clinical conditions in which diagnosis is most critically dependent upon spatial and contrast resolution. Examples include:

i) pneumothorax
ii) sub-periosteal resorption (hands)
iii) pneumocystis pneumonia (chest)
iv) breast microcalcification (mammography)
v) fine ulceration in inflammatory bowel disease (barium enema films)
vi) pulmonary interstitial shadowing (chest) - fibrosis, septal lines
vii) small pleural effusions in neonates
viii) necrotizing enterocolitis (paediatric abdomen)
xi) pulmonary nodules

Further categories were also considered, including the following:

The "almost normal" chest: a group of 100 cases of which half would contain subtle but important abnormalities of FRCR examination standard. A suitable collection of cases was available that could have been used. A particular problem at the start of our study, however, was that existing methods of handling the data made it extremely difficult to process information covering a multiplicity of diagnostic categories.

Adult ICU chest cases: these provide an opportunity to demonstrate possible technical benefits of the digital display, since the circumstances in which these examinations are made often result in films of poor quality; case material was readily available.

Casualty cases: it would be interesting to study a routine case workload. Roughly 100 patients pass through St Mary's casualty X-ray department each day, generating a total of 200-300 images. Data from Guy's Hospital (Dawood 1983) suggest that casualty officers perform 70 skull examinations for each skull fracture that is detected. To digitize a consecutive series of cases in which some of the most important diagnoses were represented just once, could therefore be a massive undertaking. (This is the problem with any attempt to apply objective methods of assessment to case material in which the incidence of significant but subtle abnormality is likely to be extremely low.) It might have been feasible to digitize 24 hours' consecutive cases, but no more; even this would have generated twice as many images as in any other category so far discussed; the majority of these cases would have been normal. The proportion of abnormalities would be small; while that would make this group more difficult to assess statistically, it would more closely simulate the everyday routine reporting workload, where the expectation of abnormality is also small.

Examination-type cases: no study can encompass every single possible difficult diagnosis. In this category, there could have been a broad mixture of 100 difficult cases of the type used in postgraduate (FRCR) radiology examinations, so that a variety of subtleties could be included. When such series were proposed, it was necessary to give prospective observers firm reassurance that no comparison would be made **between** individual observers with regard to accuracy of diagnosis, and that the primary purpose of the case material was to permit assessment for each observer of consistency of diagnosis from one image format to another rather than accuracy alone; the question to be answered was: "do any of the displayed image formats place the observer at a disadvantage with regard to film, and if so, by how much?"

Medico-legal issues: manufacturers, radiologists and clinicians are or will be aware of the medico-legal implications of digital radiography, and a further category that was considered was a group consisting largely of cases that had been the subject of medico-legal dispute. The involvement in this project of Dr Oscar Craig, who is the Radiological Advisor to the Medical Protection Society, clearly placed us in a privileged position to obtain such cases, though problems might have arisen in cases where only copy films were available.

While it is clearly essential to examine the ability of the new technology to discriminate between the presence or

absence of subtle radiological features, this is obviously an artificial situation. What really must be established before the displayed images can be brought into routine practice is that they can be used to make a complete diagnosis from scratch with no prior information about the patient (i.e. primary diagnosis) - and to make a diagnosis in difficult and borderline cases.

Several important issues are raised. It is clear that a system cannot be tested using only obvious cases; subtle features must be present, and there must be a significant chance of a feature not being visible: it is necessary to be able to create doubt in the mind of the observer. This necessarily introduces an element of bias. It is possible to use series of images that more closely represent the routine workload, but results in these circumstances would inevitably also be affected by a priori factors, such as knowledge of the types and incidence of likely radiological features.

Another issue is that, in order to evaluate such studies, the truth of the observation must be known. This implies either that the evidence for the presence or absence of a feature must be determined from external evidence, or that 'truth' must be established by inspection of the radiological images by some 'higher authority', typically a panel of experts. In the latter situation, the data is being used itself as a "gold standard", which clearly must be dangerous. This issue has been considered in the

Literature Review, but there does not seem to be an ideal solution to this dilemma. The safest approach seems to be to compare systems using well-validated sets of test images, and to perform several trials.

The reasons why some of the above clinical series were not undertaken will be considered later. It is important to recognize that it is not practically feasible to subject new systems to evaluation with every known clinical condition. Those selected for study are described below. The choice of test material has also to take account of the formidable practical constraints of collating validated clinical series of images.

All radiographs used in this work were originals of good general quality; no copy films were used.

3.2.1. Fracture series - familiarization and validation

This set of images was used at the start of the experiments, to enable the observers to become familiar with the equipment and the data collection procedure, and to monitor training.

It consisted of a small set of casualty cases - 20 cases in all, of which 10 were abnormal. These were all taken from the everyday casualty workload and showed subtle but definite fractures; in each case, the diagnosis was clearly evident on film.

3.2.2. Sub-periosteal resorption

Sub-periosteal resorption in renal osteodystrophy is a condition that has been used by other workers(e.g. Bramble et al 1987, Murphey 1989) who have studied the effects of digitization upon the diagnostic process. (Previous work has however centred largely upon comparisons between film and digitized hard copy).

A series of 40 hand radiographs was used; half the images were from patients who had no known renal disease, and who had attended the X-ray department for a variety of unrelated conditions; the other half were from patients with proven chronic renal failure in whom a diagnosis of sub-periosteal resorption had been evident radiologically on at least two occasions (before or after the image used in the study); the diagnosis was agreed by a consensus of three radiologists not participating in the study. Patients with obvious metaphyseal changes or other ancillary features were excluded. Some of the cases selected were extremely subtle.

The size of this series was restricted to 40 cases, partly because of the operating constraints and delays associated with getting System 1 up and running.

3.2.3. Pneumocystis carinii pneumonia

This second series addressed a larger group of images. We selected the subtle changes of early Pneumocystis carinii pneumonia (PCP) as suitable material.

51 chest images were selected from patients with pathologically proven PCP, with images from 52 normal patients as controls.

Since large numbers of AIDS patients are treated at St Mary's, suitable test images were relatively easy to obtain. The radiological findings of patients with PCP often lag behind their clinical symptoms, such that patients who have severe infection may have a normal or almost normal chest radiograph. At St Mary's, patients in whom this diagnosis is suspected undergo chest radiography, immediately followed by bronchoscopy with biopsy.

Pathological records were retrieved for patients who had undergone bronchoscopy for suspected PCP. Radiographs taken within 3 days of the date of bronchoscopy for patients with a confirmed positive cytological or histological diagnosis were retrieved from the film archive. Cases showing gross abnormality (e.g. effusion, localized consolidation, cavity, white-out) or ancillary features (such as oxygen masks, NG tubes etc.) were excluded. Patients in whom a concurrent diagnosis of pulmonary Kaposi's sarcoma was confirmed or suspected were

also excluded. A total of 51 suitable images were retained for study.

52 normal chest radiographs, from age- and sex-matched individuals who had undergone routine radiography in the absence of pulmonary or systemic symptoms, were retrieved using departmental computer records for use as controls. Most had undergone radiography in the Accident & Emergency Department following minor chest trauma. These images were also matched for radiographic technique (PA or AP projection, etc).

Unlike many previous image series, pathological confirmation was available for **all** of the "abnormal" test images, allowing an external, non-radiological "gold standard" to be applied. Furthermore, the conditions of the experiment were such as to enable any improvement in diagnostic performance (e.g. following contrast manipulation), relative to the original film, to become apparent.

3.2.4. Mammography

The next set of images was a series of mammograms. The objective was to study the visibility of malignant microcalcification on displayed images, since microcalcification requires high contrast and spatial resolution.

The series consisted of 62 images from patients with histologically proven carcinoma of the breast, of which 33 had been deemed to show microcalcification; a consensus on the findings had been reached by three radiologists (two of whom were specialists in mammography) viewing the original films independently of each other. All of the films used were taken prior to surgery or radiotherapy.

Malignant microcalcification is a feature more likely to be present if there is a co-existing mass, which is often more readily identifiable than the calcification itself. It was felt inappropriate to use images that did not show a mass as controls; all of the images had therefore been selected to show a mass, and the observers were instructed to make a judgement purely on the presence or absence of microcalcification.

In many cases, both the lateral oblique and the cranio-caudal views from a given patient were included. In a number of cases, microcalcification was clearly evident on one view, but not on another, even though the area of interest had been included on both views. In these cases it was agreed that the microcalcification was actually present even though it had not obviously been detected. These cases were treated as "positive", for the purposes of the experiment; contrast manipulation might arguably have been able to render the microcalcification more visible than on the original film.

It had been hoped to acquire a larger series, but the required standard of proof restricted the numbers of images that could be brought together within a reasonable time.

It is appropriate to point out that digital imaging of the breast has many theoretical advantages: image enhancement and manipulation, wide exposure latitude, reduced patient dose, and the prospect of automated image analysis, might overcome many of the limitations of present day breast imaging.

With reference to the comments of Kimme-Smith et al (1989a) already referred to (page 73) regarding the influence of observer experience, it is appropriate to point out that the radiologists with the greatest mammography experience were used in the case selection panel, and therefore could not participate as observers. All of the observers who took part were fully-trained, general radiologists, with roughly equivalent experience in mammography.

3.2.5. Skull fractures

This final series was the largest, and consisted of 286 films, 123 of which were from 43 patients with skull vault fractures.

The patients with fractures had all been referred to a specialist neurosurgical unit for treatment. Their original films had been reported as showing a fracture;

these films were also judged to show fractures by a general radiologist and a specialist neuroradiologist, working independently.

No attempt was made to select cases that were unduly subtle, and the selection procedure and need for consensus biased the case sample **away** from lesions of excessive subtlety. The study group was considered to represent a typical spectrum of radiological cases, including gross examples: there were patients with depressed fractures, intraventricular and sub-arachnoid air-fluid levels, and air-fluid levels in the sphenoid sinus.

Skull films from patients with no recorded history of skull trauma (most were patients attending a general neurological clinic), matched for age distribution, were used as normal controls. There were 163 normal films, from 53 patients.

Up to five images per patient were included. The images were not presented to observers by patient, but were viewed in a totally random sequence. This allowed the option of including material from patients in whom the fracture was obvious in one projection, but more subtle in other views.

3.3. PHYSICAL TEST MATERIAL

Although the primary concern of this study is with clinical diagnosis, interest also focused on some of the measurable physical parameters that influence image quality.

Differences in quality between the displayed images and original film do not appear to be purely a function of matrix size, and are partly the result of a multiplicity of physical factors.

In the course of this work it became necessary to develop methods, equipment, and test objects for measuring some of the physical parameters that were found to be important. This included generating test images that were also used in performance studies.

Details and results of these experiments are presented separately below.

3.4. ASSESSING AND REDUCING OBSERVER VARIABLES

All of the observers participating in the clinical experiments were radiologists of Consultant or Senior Registrar status, and were Fellows of the Royal College of Radiologists.

The individuals who took part are acknowledged on page 15.

3.4.1. Training

Prior to the start of the study, observers were shown how to use the equipment, and took part in a number of short learning exercises until they were familiar with it. These provided estimates of the likely shape of the ROC curves

and variation between observers, enabling confidence regarding the size of the image groups and the numbers of observers.

Preliminary exercises included use of the short general series of casualty cases described above.

In every subsequent pathological group, performance with a small sample of images at the start of each series was compared with performance with the same images after all of the other images had been shown, to monitor possible learning effects occurring during the study.

3.4.2. Viewing conditions

All possible identifying marks on each film that might influence an observer were concealed, and all patient names were masked with opaque material prior to digitization. Once each group of films had been accumulated in its entirety, it was digitized. Care was taken to ensure that the correct contrast range was captured by the digitization process: this was verified visually, since only limited user control of the acquisition parameters was possible.

The objective was to simulate realistic viewing conditions as closely as possible, and no undue restrictions were imposed on the observers. With System 1, the displayed images were presented on an array of 4 monitors in two rows (Figure 3.2); with System 2, the displayed images were

presented on a horizontal array of 4 monitors; an image from a separate case was displayed on each. Window width, window level, and two levels of magnification could be controlled by observers using both systems, though there were differences between the systems in terms of the precise controls and general user interface (and the over all impression was that the user interface in System 1 was preferred). Any such parameter altered by an observer was automatically reset as the next image was retrieved. A single keystroke summoned the next group of 4 images.

Ambient light was reduced to a low level, and care was taken with the equipment layout to eliminate sources of reflected glare. Ambient audible noise levels were high for both systems (System 2 was measured at 68 decibels), and little could be done to reduce this.

Frequent service and preventive maintenance was conducted, to ensure consistency of performance over the duration of the study. Quality assurance issues will be considered later.

The film images were also displayed in groups of 4, pre-loaded onto sliding frames in a viewer. Extraneous light around smaller films was masked off. Normal viewing distances were used. The observers had available a bright light and a magnifying glass for use if required, and ambient light was low.

The time allowed for the interpretation of each image (film or display) was unrestricted, and observers were permitted to change any response, prior to moving to the next group of four images, if they so wished.

Image groups were generally viewed in increasing order of likely image quality: thus, for example, a "1K" viewing session would precede a "2K" viewing session; and displayed image sessions would almost always precede any viewing of film, to reduce the likelihood of a lesion being seen at high resolution, and being remembered if the case was recognized on a subsequent occasion.

3.4.3. Randomization

With System 1, randomization of order between normal and abnormal images was achieved by modifying the inbuilt image management system to allocate each film a random number on entry into the database, prior to digitization. For each viewing session, the images were subsequently retrieved in numerical order.

No such facility could be introduced with System 2. Random number tables were therefore used to allocate each film a unique reference number, and the films were then sorted into numerical order prior to digitization.

For each image series, this order remained constant throughout. It was regarded as neither feasible nor

worthwhile to attempt fresh randomization for each viewing session and each observer.

3.5. DATA CAPTURE

For the experiments with clinical images, each observer was shown both the displayed image, and the original film, separated by a period of time ranging from not less than 2 weeks up to several months; the important comparison to be made was between the performance of each observer on film and his or her own performance on the display with the same cases.

The observers were asked to decide simply between the PRESENCE or ABSENCE of the condition under study in each case, and they recorded their observations directly onto a microcomputer with a mouse, selecting a value on a visual analogue certainty scale on which 5 points were marked, in a manner that has been well-described in the literature:

1. definitely normal
2. probably normal
3. equivocal
4. probably abnormal
5. definitely abnormal

Although there are commercially available computer programmes designed for data capture in such situations (e.g. FEASIBLE, BAZIS Ltd, Leiden, Netherlands), we

preferred the greater flexibility of a specially written programme, which was kindly prepared by Andrew Todd-Pokropek at University College London. The data collected was therefore recorded automatically in a form that could later be used for ROC analysis.

Further assessment of the optimal use of rating scales in ROC analysis was conducted as a separate methodological experiment, described below.

Use of the specially written data capture programme to input the observations enabled viewing time per image, and duration of each viewing session, to be recorded.

For each case series, observers were provided with written instructions, an example of which is shown in Fig 3.6.

Figure 3.6: Sample observers' instruction sheet.

INSTRUCTIONS TO OBSERVERS

You are about to examine a carefully selected series of chest images. Some are taken from symptomatic patients with a pathologically proven diagnosis of Pneumocystis carinii pneumonia PCP). Others are normal examinations, matched for age, sex, and radiographic technique, from patients with no clinical features of infection; in the majority of cases, the images in the "normal" group were taken in the casualty department following minor chest trauma.

A small number of cases showing gross abnormality has been excluded from the PCP group. Examples of features justifying exclusion include: effusion, localized consolidation, cavity, white-out, visible oxygen mask, ECG leads, tubing etc.

Use any of the features on the images presented to you, to decide whether or not you think that PCP is the diagnosis. The proportion of normal cases lies somewhere between 30 and 70 per cent.

Do not feel obliged to make a definite decision on normality or abnormality (points "1" and "5" on the certainty scale); feel free to use any point on the scale, not just the whole numbers.

The purpose of the study is to compare detectability of abnormal features on displayed images versus the detectability of the same features on film.

3.6. DATA HANDLING

Receiver Operating Characteristic (ROC) analysis is the most widely accepted objective means of evaluating diagnostic performance. It is important to note that it places full emphasis on diagnosis - the end result - and frees us, to an extent, from having to evaluate individual physical or psychophysical determinants of image quality such as spatial or contrast resolution. Results from ROC studies are usually presented in the form of ROC curves - a plot of True Positive rate against False Positive rate. A simple explanation of the ROC curve is presented in Appendix 2.

The intended use of ROC analysis influenced the choice and numbers of cases, and the structure of the experiments.

The data capture programme presented the results of the observations in a form that would allow rapid transfer to ROC analysis software such as ROCFIT, and this will be described in greater detail under "Results".

The results of the observations were evaluated using ROC analysis and ROC curves.

4. METHODOLOGICAL EXPERIMENTS

4.1. INTRODUCTION

In the course of conducting the observation experiments, it became clear that further issues relating to the manner in which data should be handled justified attention. Experiments were conducted to address the following issues:

- The effect of using continuous rating scales for data capture, rather than discrete scales.

- The frequency with which different points on the rating scales were used, and the effects of attempting to achieve more uniform use of the rating scale.

- The effects of using pooled data in ROC experiments.

4.2. THE USE OF CONTINUOUS RATING SCALES IN ROC ANALYSIS

Observers participating in ROC studies are usually required to estimate the confidence with which each observation is made. With a discrete scale, the rating or score normally falls into one of 5 categories, ranging from 'definitely normal' to 'definitely abnormal'.

A problem exists, however, concerning the separation of points on an ROC curve: in general, the more points that are used on the rating scale, the better separation it will be possible to generate for the points on the curve; for a small number of observations, the points are often badly separated. The typical 5 rating experiment generates only 4 points on an ROC curve, and it is not uncommon for ROC experiments to yield curves where there are 3 points on the y-axis, and only one point in the middle of the ROC plot space. Such data simply cannot be fitted to a curve, and is an example of the so-called degenerate case referred to later in this section (Figure 4.4, page 129). Degenerate cases occur when there are no differences in the number of false positives for several different ratings. A frequent example occurs when there are NO false positives for ratings 3 to 5. In such cases, the ROC data is unusable.

One approach to this problem is the use of a continuous visual analogue rating scale, on which observers select a point using a mouse; this was the method adopted throughout this investigation (as described on page 113). Using a continuous rating scale might be expected to improve the ability to generate well separated points on the ROC curve, and to decrease the likelihood of the 'degenerate' case.

The following experiment was conducted to explore this hypothesis. A set of observations (in fact, the observations recorded with the hand images viewed on System

2) was rescored, and the value for each response was converted to the closest integer. The two sets of data were compared, using CLABROC (the continuous rating correlated data analysis programme from the ROCFIT software package) for the continuous data, and LABROC (also part of the ROCFIT package) for the data that had been rendered discrete. The ROCFIT software package is published by Metz (1987).

It was possible to demonstrate only a minor statistical difference was between the two sets of data. However, the use of a continuous scale is still to be preferred on theoretical grounds, since the likelihood is that this will render tests statistically more sensitive.

4.3. ACHIEVING MORE UNIFORM USE OF RATING SCALE CATEGORIES

Radiologists are trained to be decisive, and aspire to reaching a conclusion regarding their interpretation of each image: a major problem in data analysis from ROC studies is caused by observers who have not used the points on the rating scale in a uniform manner, and have made many responses corresponding to the two extreme categories with few responses falling in the middle.

To solve this problem, it has been suggested that observers should be trained to use the rating scale in a more convenient manner for the subsequent data analysis. However, forcing observers to function in an unnatural manner might lead to impairment of performance, and the consequences require investigation.

4.3.1. Method

A group of observers was asked to re-score a set of difficult clinical images after re-training. The clinical material used for this experiment consisted of the group of hand radiographs from patients with subperiosteal resorption, that had been used with System 1 and in the comparison between System 1 and System 2.

On the first occasion that these images were viewed on System 2, the observers had been allowed to report the images as they pleased, with no special instructions regarding the rating scale. A continuous scale was used, and observers recorded their findings by selecting a point on the scale with a mouse. It was observed on that occasion, that a total of 56% of all reports had fallen into the two extreme categories.

After an interval of 10 months, the same images were re-examined by the seven observers at that time available (of the original eight); on this occasion, however, the observers were instructed to attempt to use the rating

scale in as uniform a manner as possible. To reinforce this request, a continuously updated histogram was displayed, showing how they had used the rating scale up to that point (Figure 4.1), and providing continuous feedback on their compliance. The observers were also told that exactly half of the images were positive. At the same time, a repeat trial for a limited set of images was conducted to confirm reproducibility.

4.3.2. Results

The number of reports falling into the two extreme categories was reduced, but only to 32% of the total. The ROC curves from the two studies were compared using CLABROC.

The distribution of the points used on the rating scale during the two viewing sessions is shown in Figure 4.2.

The corresponding ROC curves are plotted in Figure 4.3. The areas under the two curves were 0.8598 for the original data, and 0.8178 for the 'uniform' data, which were not significantly different (p=12%).

In the original trial, the repeat experiments to check reproducibility showed much smaller differences than those observed between the two viewing sessions in this experiment.

Figure 4.1: Observers' data entry screen, modified to display a continuously updated histogram of the response distribution. Observers use a "mouse" to select a point on the rating scale in the lower portion of the screen. The boxes in the upper left portion of the screen represent the bank of monitors, and a highlight in one of the boxes serves as a reminder of the location of the currently relevant image.

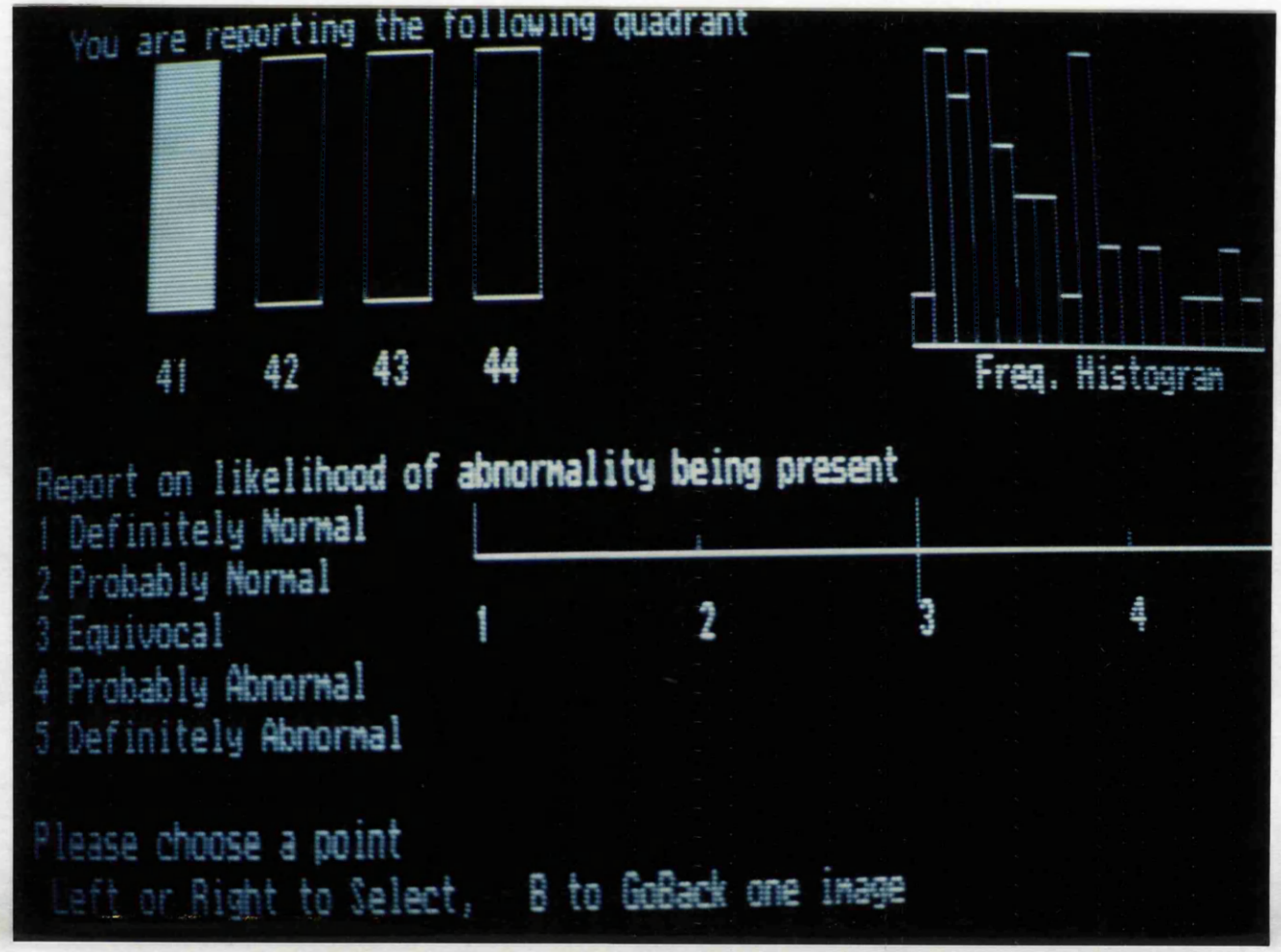
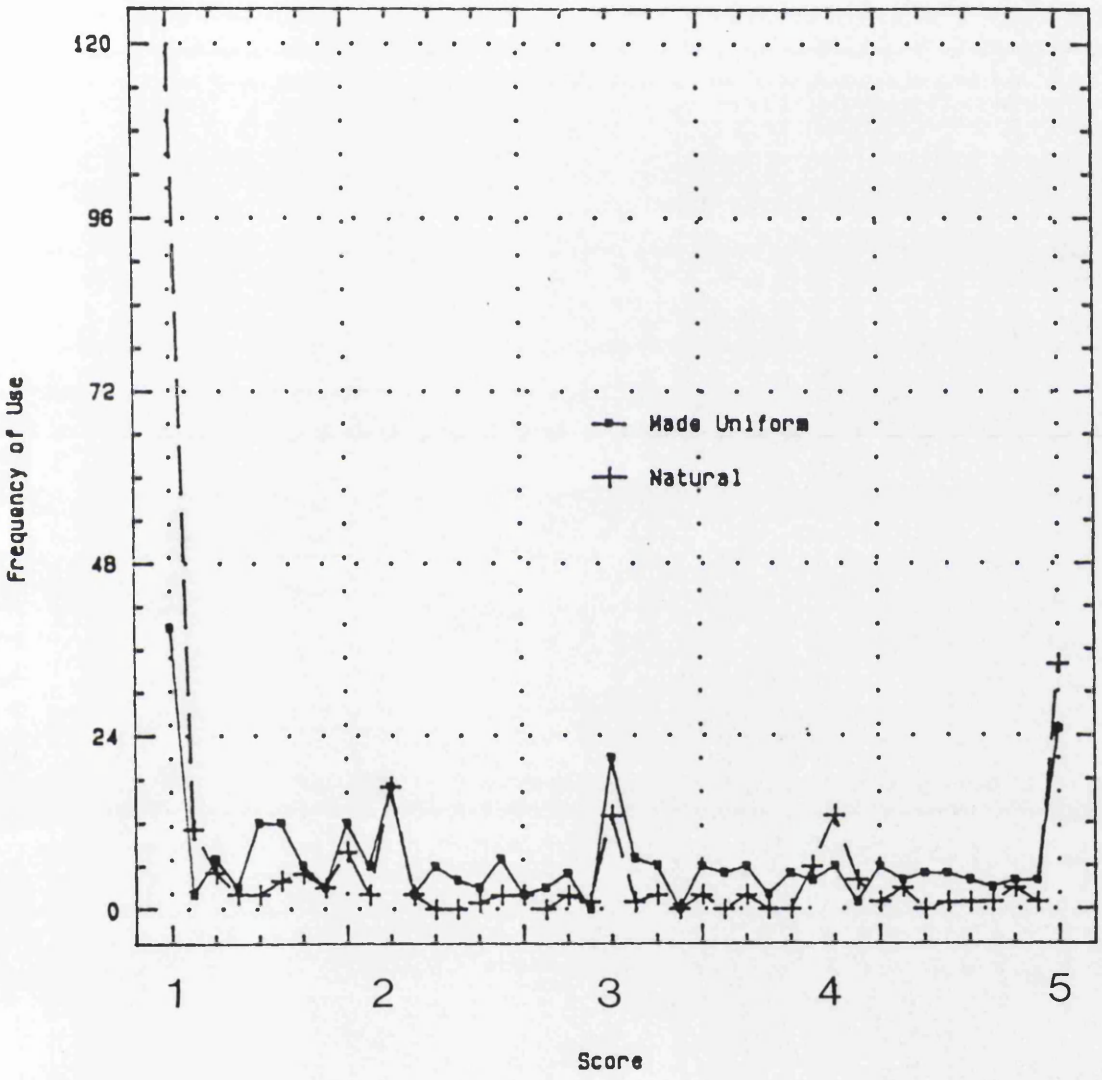
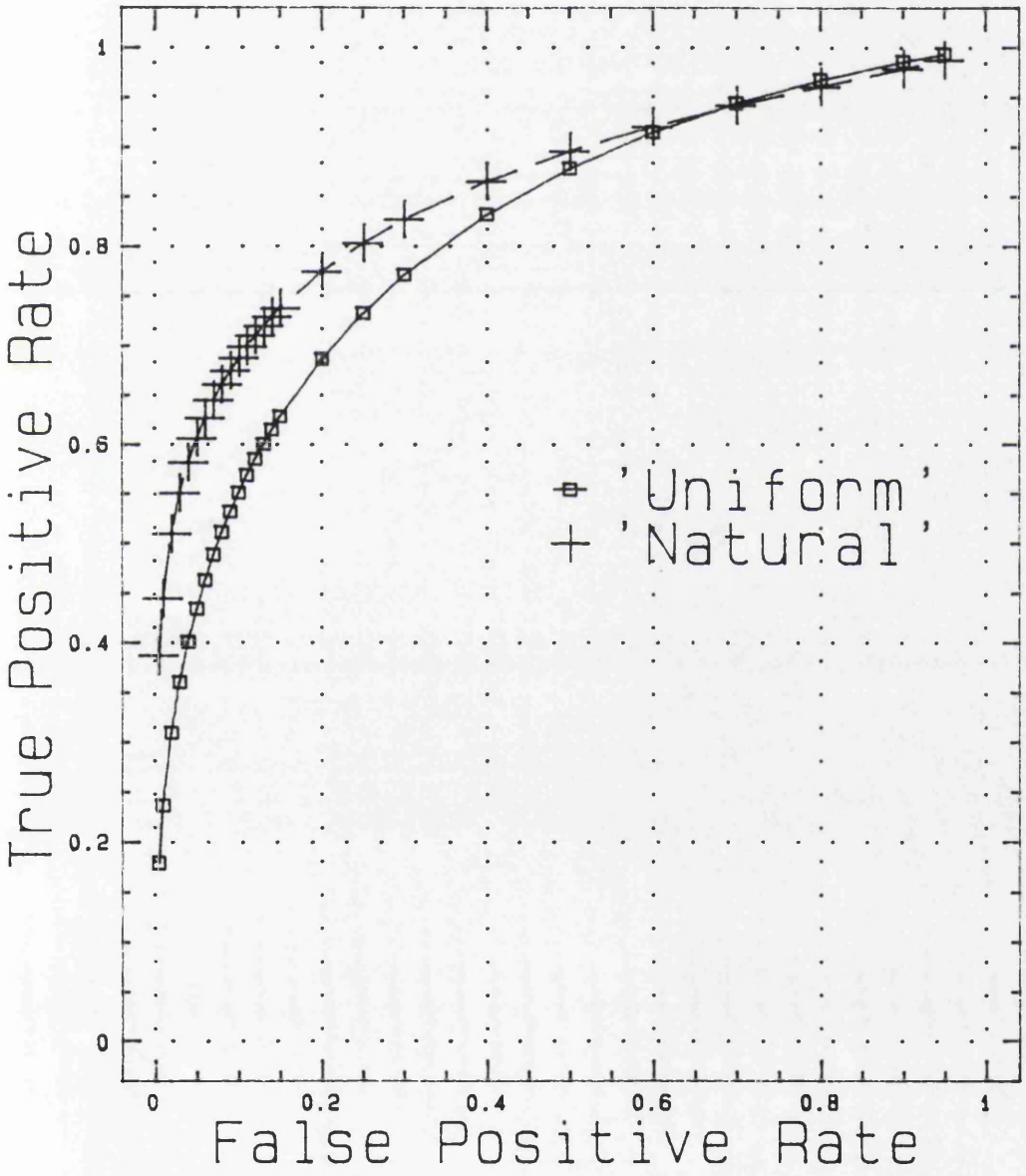


Figure 4.2: Distribution of the points used on the rating scale. "Natural" refers to the first viewing session; "uniform" refers to the second, in which observers were encouraged to achieve a more uniform distribution of responses.



Note that the histogram is 'peaky'- the observers tend to use preferentially the conventional categories, and secondly, that there are enormous spikes for the categories 1 and 5. In fact, in this experiment and typical of many others, 70% of all observations were scores as either 1 or 5, and only 30% fell into the range 1.1 up to 4.9.

Figure 4.3: R O C curves, with and without encouragement to achieve a uniform distribution of responses. The areas under the two curves were 0.8598 for the "natural" data, and 0.8178 for the "uniform" data, which were not significantly different (p=12%).



4.3.3. Discussion

The purpose of this experiment was to assess what changes such retraining might cause. If it introduced "noise" in the observer, it might be expected that the ROC curve following retraining should fall.

This is precisely what was observed. However, the difference in the ROC curves before and after retraining was not statistically significant.

It should also be noted that all ROC experiments such as those described here, with or without re-training, represent an artificial situation that does not truly simulate the real-life reporting tasks that radiologists undertake in clinical practice.

The loss of diagnostic efficiency in this study was not statistically significant, but this probably reflects the difficulty observers found in trying to make their responses more uniform, since the histograms of rating scale usage were also not very different.

4.3.4. Conclusion

The re-training of observers to use the rating scale in a more uniform but artificial manner, can introduce bias,

which tends to lower resulting ROC curves, and should be avoided.

4.4. SHOULD ROC DATA BE POOLED OR PAIRED ?

The statistical analysis of results from evaluation of different diagnostic display systems is usually carried out using ROC methodology. It was necessary to consider the question of whether such an analysis, when using multiple observers, should be carried out by pooling the observers' data, or by pairing the observations for the two systems under comparison. This work was largely conducted by Andrew Todd-Pokropek but is described here because it arose from combined efforts to find the best way to handle data arising from the experiments described above, and has implications for the interpretation of results.

The way that most image quality investigations take place is that observers are asked to evaluate two (or more) sets of images, and for each image within each set provide their interpretation together with a confidence rating, from which ROC curves can be constructed.

It is important to recognize that such data is correlated. It is possible to generate ROC curves by a suitable bi-normal fitting process in probability space, and then to compare the resulting curves, but in order to state that System A is better or worse than System B there must also

be an associated assessment of the confidence with which that statement can be made; in other words: is the difference between the two curves significant?

When analyzing ROC data - especially with a relatively small number of observations such as occurs when using un-pooled data - a frequent problem is that curve fitting fails as a result of badly-placed points in ROC space (Figure 4.4). In particular, this happens when several of the points lie on the y-axis (they have a zero false positive rate), which often occurs when observers use the rating scale in a conservative manner.

An assessment of the correlation between the data is critical when trying to assess the significance of the differences between ROC curves. It must also be pointed out that such data can alternatively be analyzed by methods other than by fitting a maximum likelihood line in probability space, for example by non-parametric methods, which have less restricting assumptions.

The key issue is the effect of inter and intra observer variation. Figure 4.5 shows one typical ROC experiment, where the ROC plots are shown for 4 observers when reading film. It may be noted that the variability between observers is considerable.

Figure 4.4: R O C points in the "degenerate" case: curve fitting fails when several of the points lie on the y-axis.

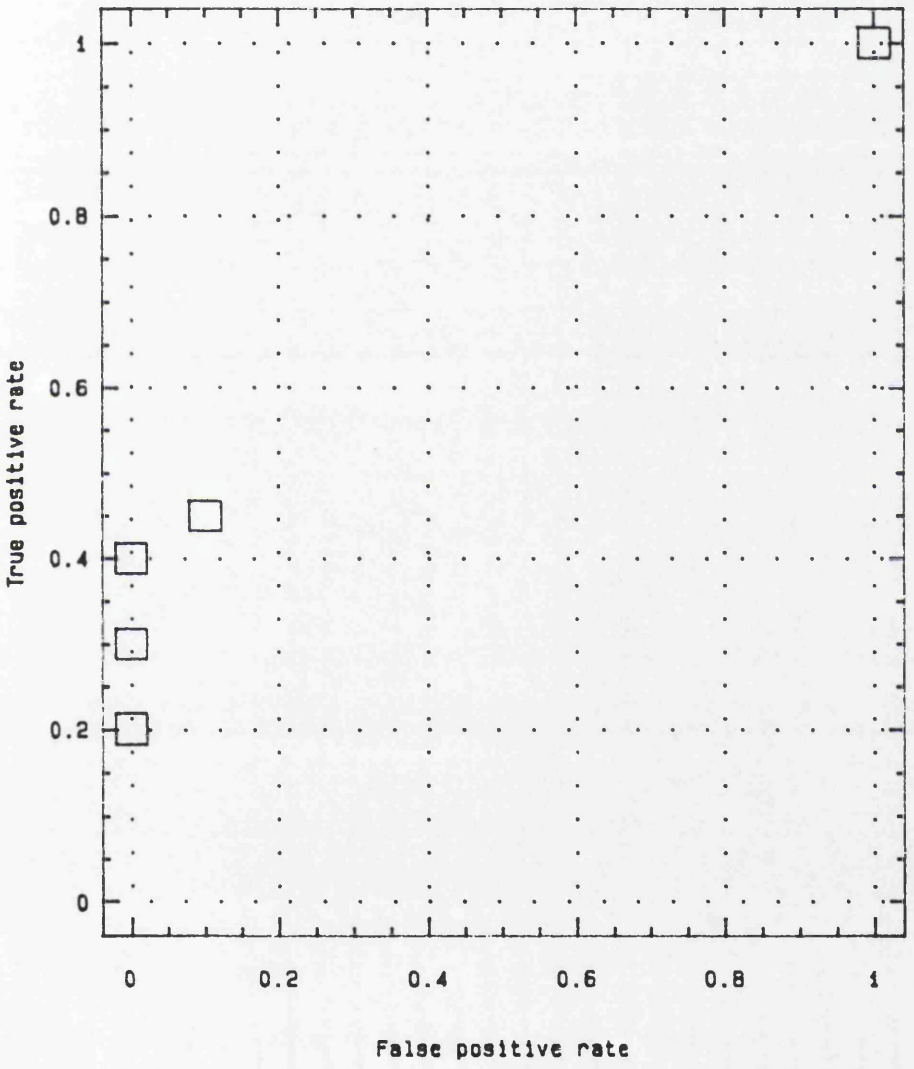
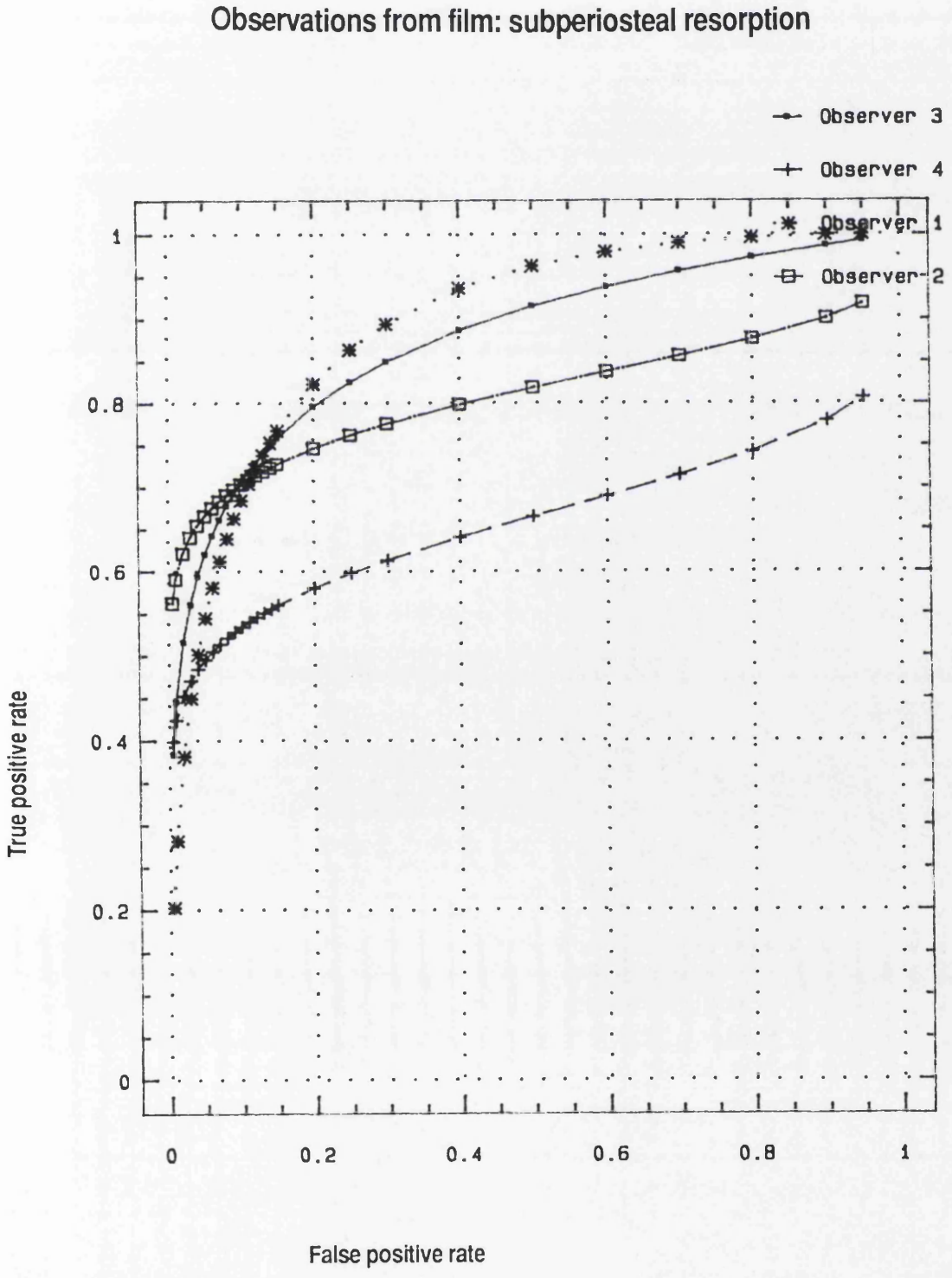


Figure 4.5: R O C curves in an experiment with four different observers. This illustrates inter-observer variability, and that differences may be at least as large as the differences between the different 'systems' being compared.



The usual method of assessing such data is to pool the results for all observers for each 'system' being compared. Such a method loses all information about individual observers and therefore is, strictly, wrong.

The main reasons for pooling such data are to increase the number of observations from which each ROC curve is generated, and to reduce the effect of observer variability. One of the effects of pooling is to pull the ROC curve down. However, this will occur for each 'system' being compared, and therefore may not be important provided that the differences between the two curves and the estimation of the significance of the difference are unaffected.

When paired data is used (that is, an analysis of the significance of the difference is made for each observer separately) the inter-observer variability is taken into account directly. As for all paired tests, the statistical power should increase. However, for an individual pair of curves, there are few observations per curve, and there is little 'sensitivity' for individual observers; the final results need to be formed by pooling the results for all individual observers AFTER the ROC analysis. The effects of inter-observer variability are excluded.

It is certainly desirable to pair data from individual observers, and then pool the results from the estimate of

the significance of differences for each observer. However, such a method creates cases where there are few observations, and often results in 'degenerate' cases.

It is important to maximize the number of observations from which an ROC curve is generated: when an ROC curve is fitted to data from few observations, the fit is poor, and the estimates of error are large. The estimate of the significance of the difference between two ROC curves will improve with increasing numbers of observations and for example will be quite poor (i.e. insensitive) for a single observer. However, increasing the statistical power of such an analysis means not just reducing the error estimates of ROC curves, but eliminating bias. The inter-observer variation which is included by pooling is such a source of bias.

Using such a method (where possible) on a whole series of ROC experiments it was found that the direction of change did not change, and the statistical significance of the difference increased.

In general, estimates of the significance of such differences are highly sensitive to the method used. This sensitivity results from the difficulty in estimating correlation. For example, it was observed that the significance decreased, and the sensitivity of the method sometimes increased (but did not get worse) both when pairing the data and when using continuous rating data.

As previously mentioned, alternatives exist, such as the use of non-parametric methods. Non-parametric methods would be expected to have much less statistical power, but there is evidence that they perform as well as parametric methods in certain cases. There is also a need for developing bi-variate paired tests.

In summary, when a series of observers looks at two sets of images, this is a paired experiment. Although one should ideally always pair data rather than pool data, the fragility of the ROC fitting process may exclude this. It is for this reason that almost all results presented from such experiments are based upon pooled data. However, there is a need for slight caution with estimates of the significance of differences in such cases.

The data analyzed here was pooled between observers. This also introduces a source of bias, since inter-observer variability is eliminated from consideration. However, it is likely that the bias introduced would have tended to reduce the significance of the differences between the ROC curves, and that, if this effect was taken into consideration the results reported here would probably be slightly more significant.

5. CLINICAL EXPERIMENTS

5.1. SYSTEM 1

5.1.1. Fracture series

At the very start of the study, as the first experiment to be performed on System 1, this series was digitized and shown to a team of eight observers as a training exercise. They were asked to adjudicate merely on the presence or absence of a fracture.

For ease of image management, the images were shown at "2K" with compression (see below). After the renal osteodystrophy series had been completed, the fracture series was shown again, to document any improvement in performance that could be attributed to learning. The observers were not shown the original films from these cases.

5.1.2. Sub-periosteal resorption

The original hand radiographs were digitized at "1K" and "2K". With the System 1 digitizer, unlike others, the precise digitization parameters varied with film size: only 2 film sizes were used in this series, and the corresponding parameters are given in Table 3.

TABLE 3: Film digitization parameters, System 1: Source: manufacturers' data.

	"1K"	"2K"
Digitization parameters for film size 18cm x 24cm:		
Area scanned	17.5cm x 22.8cm	
Pixels read	2000 x 2600	
Pixels stored	750 x 975	1500 x 1950
Pixel size	0.2333mm	0.117mm
Digitization parameters for film size 24cm x 30cm:		
Area scanned	24.7cm x 28.8cm	
Pixels read	1800 x 2270	
Pixels stored	752 x 908	1504 x 1816
Pixel size	0.328mm	0.164mm

At each viewing session, an observer would see the complete image series for a given format.

To assess consistency of response, the first viewing session of renal osteodystrophy cases was shown again to all 8 observers after all of the viewing sessions for this condition had been completed. (This was in fact again the "2K" compressed series.)

Each observer was required to arbitrate merely on the presence or absence of sub-periosteal resorption.

5.1.3. Data compression experiments

Besides investigating the effect of matrix size on diagnosis we also had an opportunity to examine the effect of data compression.

The built in DCT algorithm enabled typical irreversible compression ratios of 15:1. After a subjective assessment of the images from this series presented at "1K" with compression, it was decided that these did not merit further formal study; in particular, block artifacts were prominent and resulted in considerable image degradation. It was decided, however, to proceed with consideration of the "2K" compressed images in our evaluation.

5.2. SYSTEM 2

5.2.1. Sub-periosteal resorption

The installation of System 2 afforded an unprecedented opportunity to compare two different, commercially available systems, using the same clinical material and the same observers (in fact only seven of the original team of eight observers were still available to the project).

The original films were digitized at 210μm, corresponding to a nominal matrix size of "1K", that is, over a film size of 24cm x 30cm to a matrix size of 1200 x 1500.

These images were presented to observers in a single session.

5.2.2. *Pneumocystis carinii* pneumonia

The original chest radiographs were digitized at 210μm, corresponding to a matrix of 1700 x 2000.

The images were presented in turn to a team of nine observers. The size of this series meant that it was not always possible for observers to view the complete series of images in a single session. Each observer was required to arbitrate only on the presence or absence of abnormality suggestive of PCP.

Ten of the displayed images were shown to observers both at the beginning and at the end of the observation sessions, to assess the effect of increasing familiarity with the equipment or the condition under study.

5.2.3. PCP: with data compression

The original chest radiographs were digitized at 210μm, corresponding to a matrix of 1700 x 2000. The inbuilt compression algorithm (level 1 - see page 95) was applied to the images. Average compression ratios of 8:1 were achieved.

The compressed images were presented to the same team of nine observers, prior to viewing the non-compressed images.

5.2.4. Mammography

The original radiographs were digitized at 210μm.

The images were presented in turn to a team of eight observers, in almost all cases at a single session. Each observer was required to arbitrate only on the presence or absence of microcalcification.

Ten of the displayed images were shown to observers both at the beginning and at the end of the observation sessions, to assess once again the effect of increasing familiarity with the equipment or the condition under study.

5.2.5. Mammography: data compression

The original radiographs were digitized at 210μm, and the inbuilt compression algorithm was applied to the images, again at level 1. Average compression ratios of 7:1 were achieved.

The compressed images were presented to the same team of eight observers prior to viewing the non-compressed images.

Again, ten of the displayed images were shown to observers both at the beginning and at the end of the observation sessions, to assess the effect of increasing familiarity with the equipment or the condition under study.

5.2.6. Skull fractures

The original skull radiographs were digitized at 210μm. The images were presented in turn to a team of seven observers. The large size of this series meant that it was not possible for observers to view the complete series of images in a single session, and for practical reasons most of the displayed images had to be shown in groups of 40. Viewing sessions spanned several weeks. Each observer was required to arbitrate only on the presence or absence of a fracture.

Again, ten of the displayed images were shown to observers both at the beginning and at the end of the observation

sessions, to assess the effect of increasing familiarity with the equipment or the condition under study.

No data compression was applied to this series.

6. RESULTS FROM CLINICAL EXPERIMENTS

6.1. INTRODUCTION

The results of the observations were evaluated by constructing ROC curves of the pooled data, and the significance of the differences between them was assessed using paired analysis of the parameters of the curves assuming a bi-variate normal model (Swets & Pickett, 1986; Metz, 1978, 1986, 1988), using published ROCFIT software from the University of Chicago (Metz, 1987).

6.2. SYSTEM 1

6.2.1. Fracture series

Figure 6.1 shows the ROC curves for the first and second occasions that these images were shown. The curves are virtually identical (areas under the curves were 0.8721 and 0.8769), and there is no statistically significant difference between them.

Out of interest, it is worth noting that several important lesions were missed (bearing in mind that no comparison was made during the study between the displayed images and the original films). Among them, a fifth metacarpal fracture that had been obvious on film was missed on 8 out of 16 occasions (Figure 6.2), and an extensive parieto-temporal skull fracture, again obvious on film, was missed on four occasions.

Figure **6.1:** R O C curves for fracture images: first and second viewing.

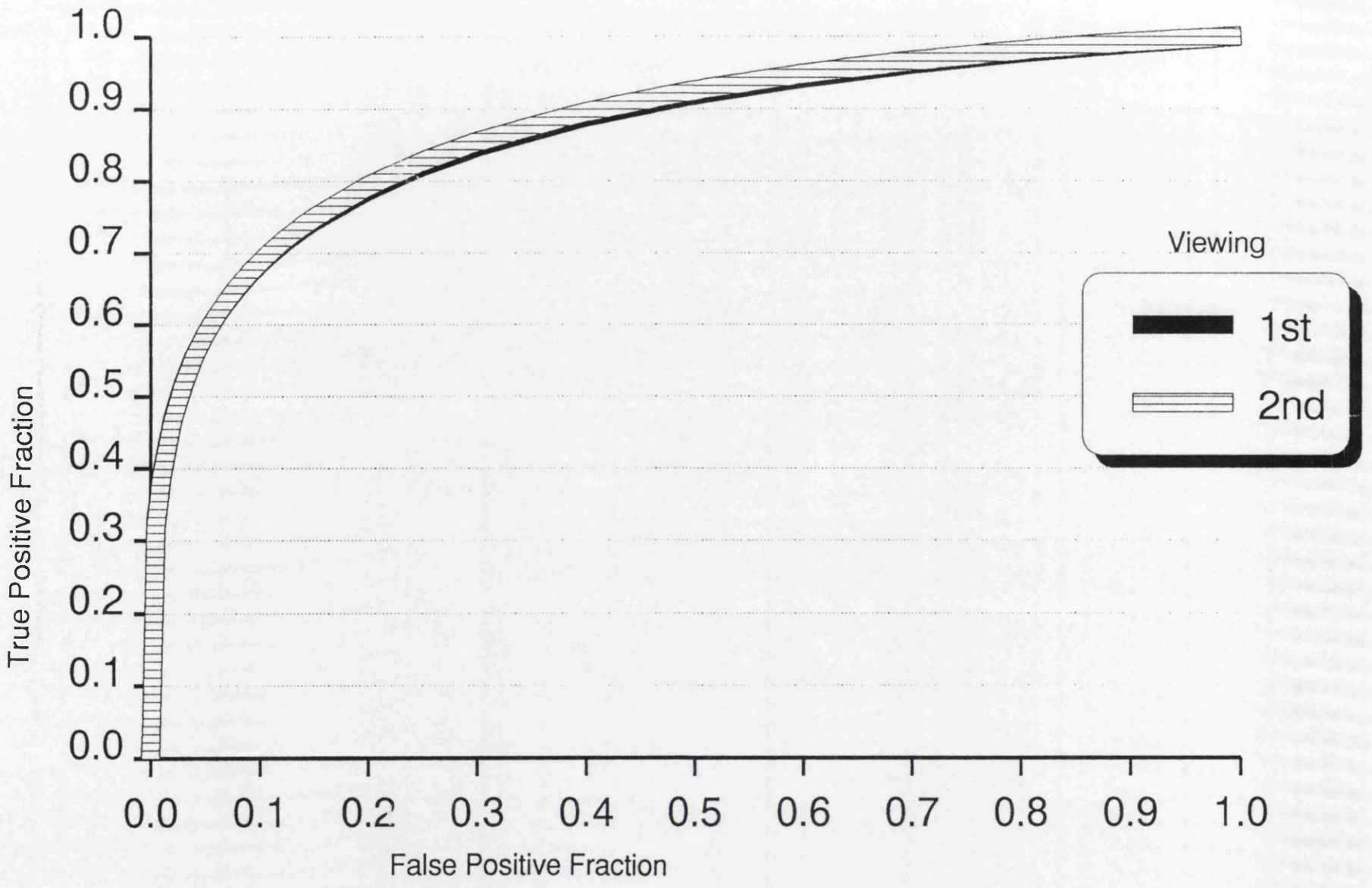


Figure 6.2: fifth metacarpal fracture, missed on 8 out of 16 occasions: display.

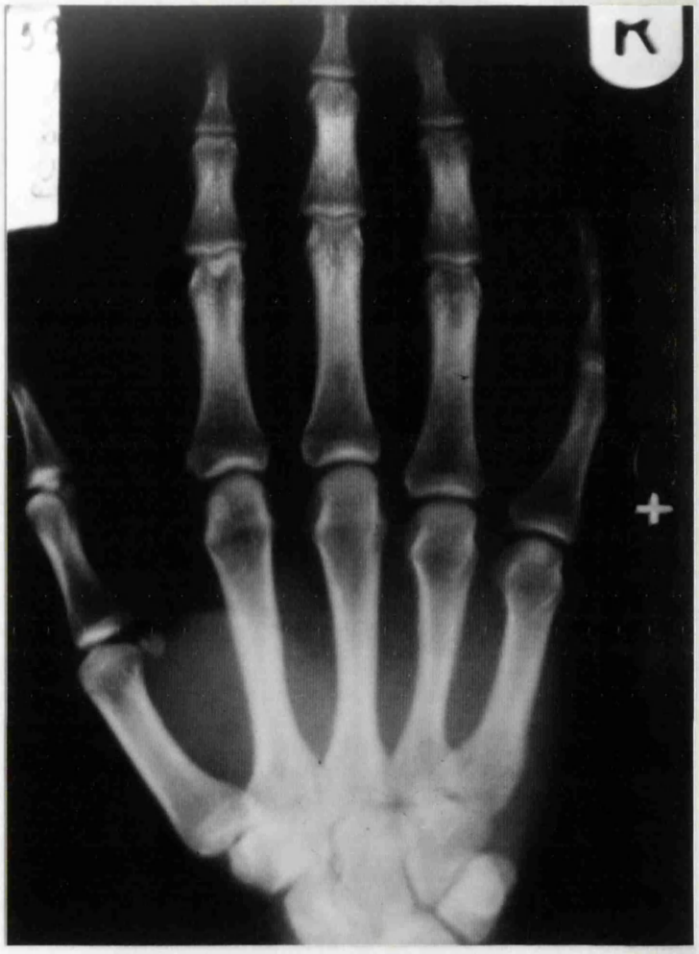


Figure 6.3: renal osteodystrophy images: R O C curves for first and second viewing.

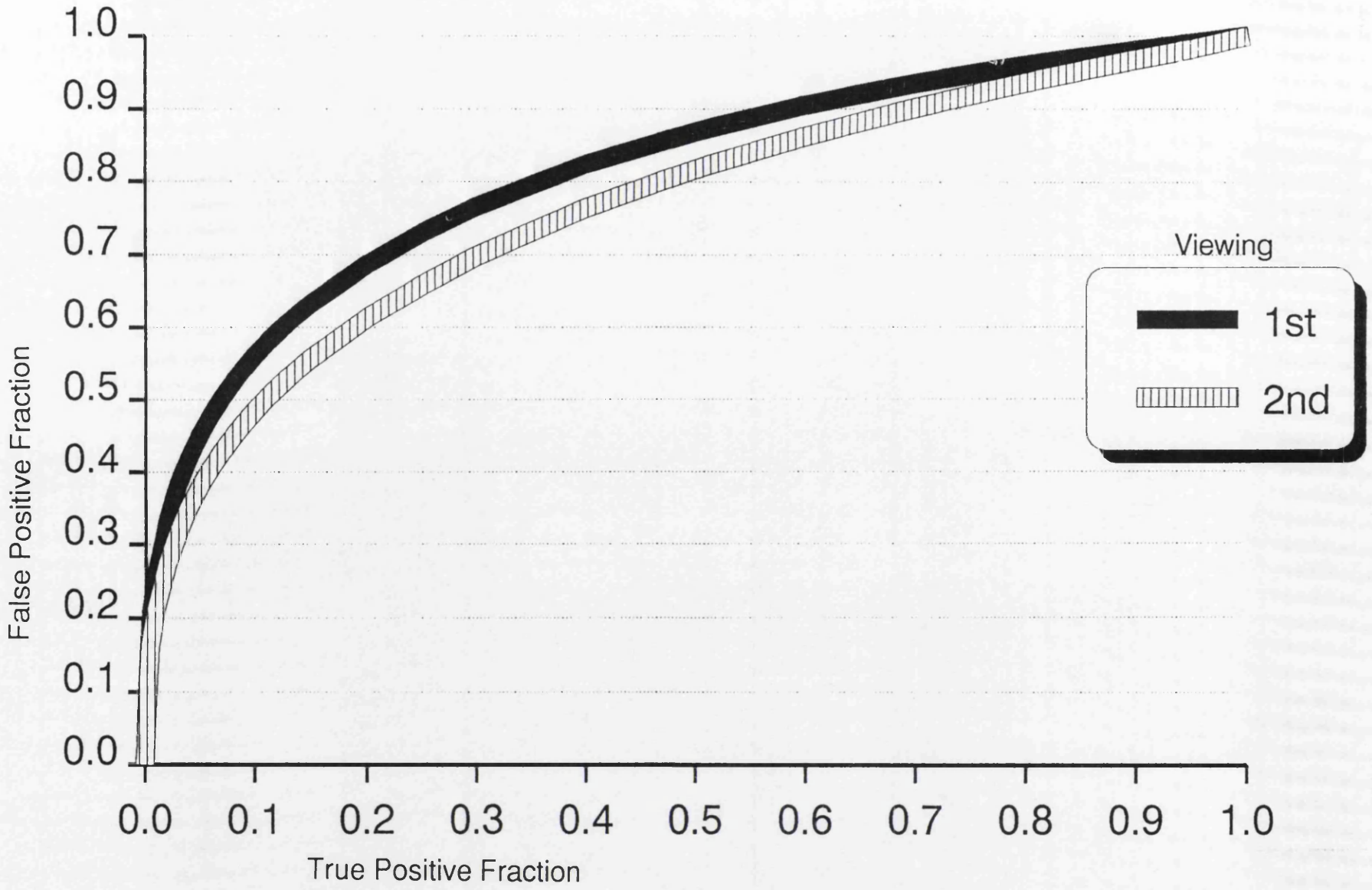
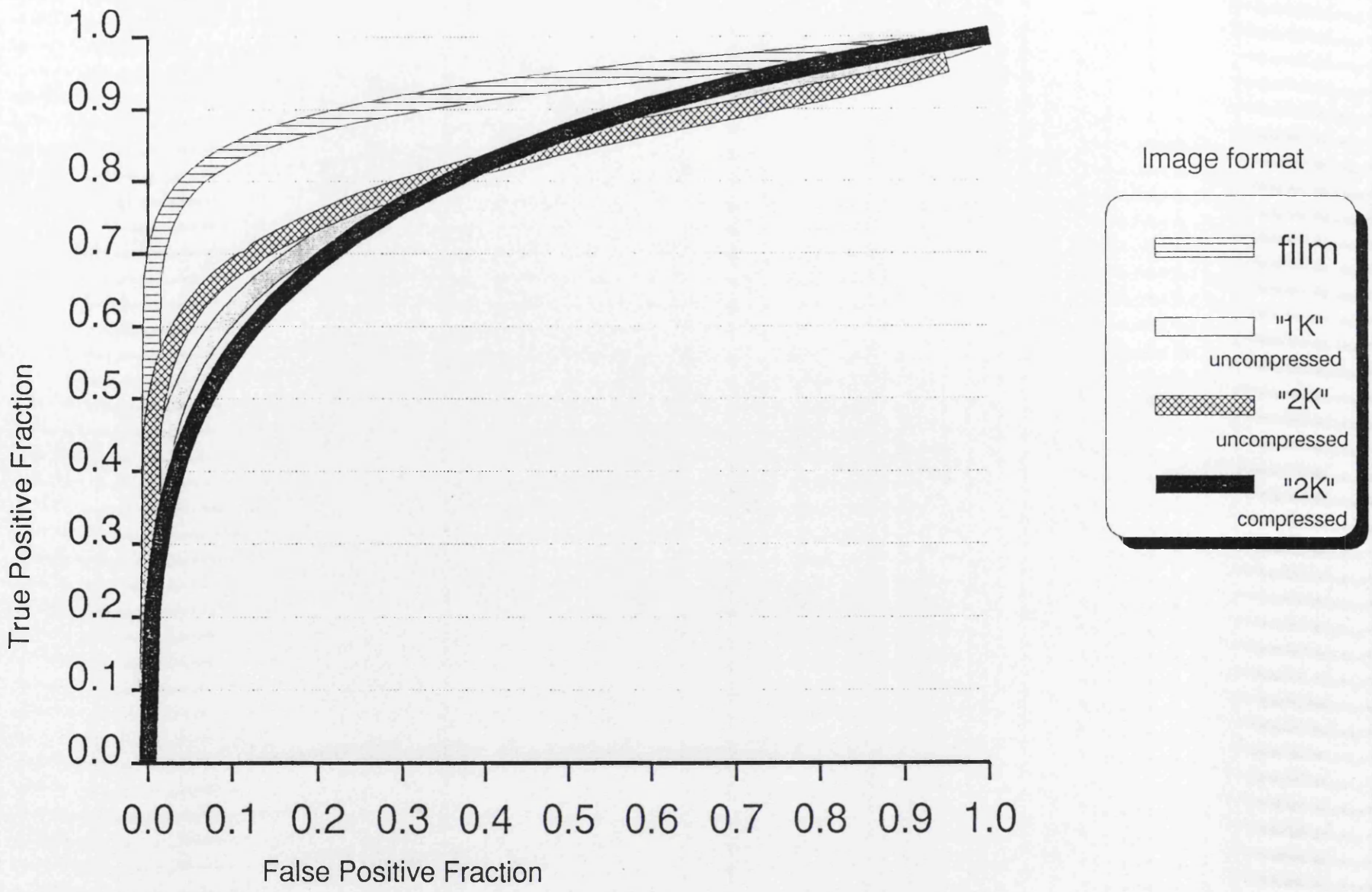


Figure 6.4: renal osteodystrophy images: R O C curves for film and display.



6.2.2. Renal osteodystrophy

Figure 6.3 shows the curves for the images that were shown twice. The areas under the curves were 0.8162 and 0.7609. The slight difference between the two curves is not statistically significant.

This result, and that above, suggest a high degree of consistency and reproducibility between the first and second viewing sessions for both series.

Figure 6.4 shows the ROC curves for the main study. The areas under the curves and the significance of the differences from film, are given in Table 4.

In order to provide a simpler means of interpretation of ROC data, it is useful to select an acceptable false positive value at which the true positive values can be compared. In presenting the results from these investigations, an arbitrary choice of a false positive rate of 15% has been made. More recently, Cox et al (1990) have also opted for such a means of summarizing the data, though they chose a value of 18.5 per cent.

Table 5 shows the true positive fraction at a false positive rate of 15%.

TABLE 5: Areas under curves, and significance of differences from film: ROC data, System 1

	Area under curve	Standard error	Significance of difference from film
Film	0.9232	0.0247	
"1K"	0.8231	0.0266	p<.002
"2K" compressed	0.8162	0.0274	p<.002
"2K" compressed (2nd viewing)	0.7609	0.0316	p<.001
"2K"	0.8299	0.0298	p<.005

There is a significant difference between the curve for film and for all the VDU displayed formats ($p<.005$).

We were unable to demonstrate a significant difference between the "1K" and "2K" images.

Examples of some of the images are presented on the following pages. The purpose of these illustrations is simply to demonstrate the kind of material that was used; it should be noted that the loss of quality in photography is in many instances comparable with the image degradation in digitization and display that this study was attempting to assess.

Figure 6.5 shows images from a patient with chronic renal failure. Images shown are at "1K" and "2K" compressed, at "2K" without compression and on film. Note the raster lines.

There is a significant difference between the curve for "2K" images, and the curve for "2K" compressed images using this algorithm ($p<.02$).

There is significant loss of diagnostic image quality on compression of the "2K" images using this algorithm, in comparison with original film.

We were unable to demonstrate a significant difference between the "1K" uncompressed and the "2K" compressed images.

6.2.3. Data compression experiments

The curve for "2K" images with data compression is also shown in Fig 6.4, and another example of one of the images is shown in Figure 6.6.

TABLE 5: True positive fractions at a false positive rate of 15%: ROC data, System 1

	TRUE POSITIVE RATE AT FALSE POSITIVE RATE = 15%
Original film	86.4%
"1K"	67.6%
"2K" compressed	63.2%
"2K"	72.1%

TABLE 6: True positive fractions at a false positive rate of 15%: ROC data, System 1 vs System 2

	TRUE POSITIVE RATE AT FALSE POSITIVE RATE = 15%
Original film	89.4%
System 1: "1K"	69.4%
System 1: "2K"	75.0%
System 2: "1K"	77.3%

TABLE 7: Differences between ROC curves: System 1 vs System 2

	Original film	System 1 "1K"	System 1 "2K"
System 1: "1K"	p=<0.002		
System 1: "2K"	p=<0.005	p=<0.85 [NS]	
System 2: "1K"	p=<0.05	p=<0.1 [NS]	p=<0.1 [NS]

Figure 6.5: images from a patient with chronic renal failure: film and display.
a) "1K"; b) "2K" compressed; c) "2K" without compression; d) system 2 ("1K").

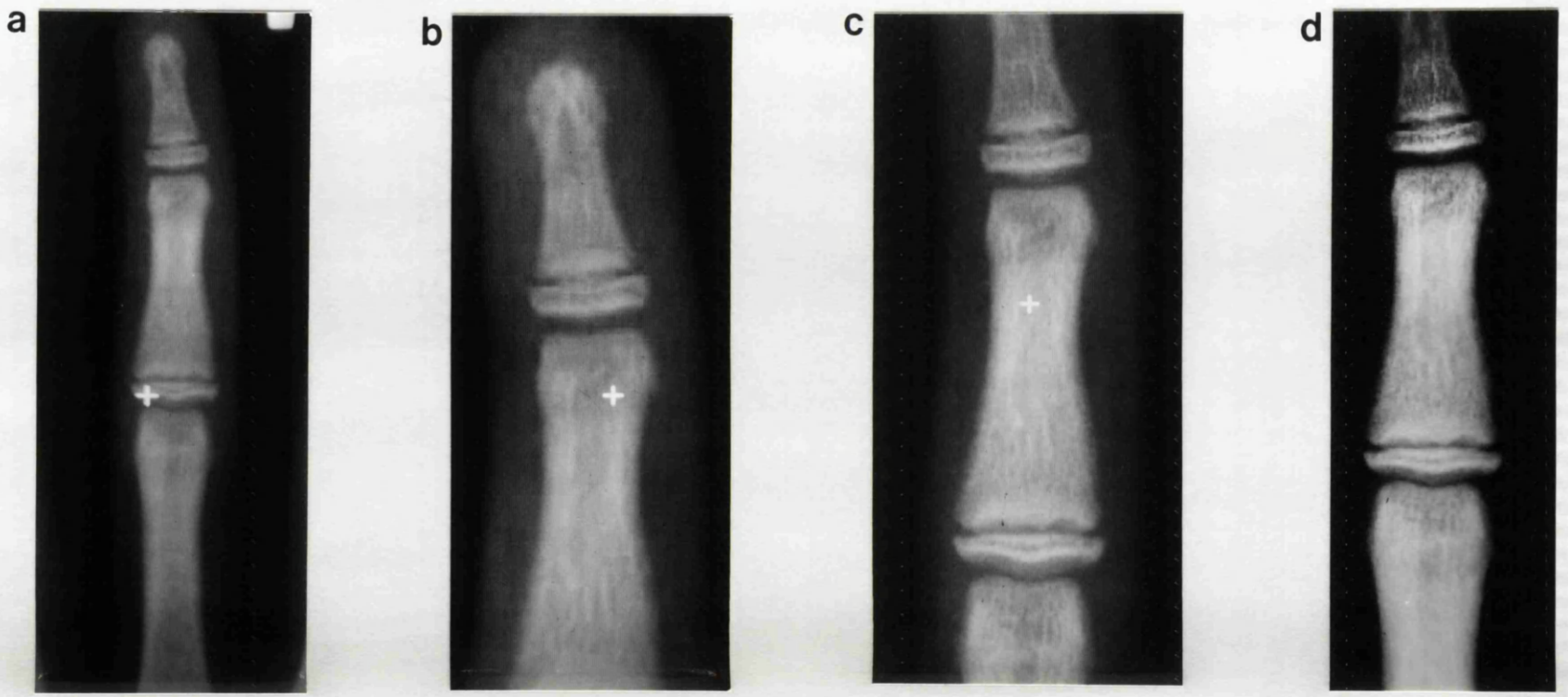


Figure 6.5 (continued): images from a patient with chronic renal failure: e) film

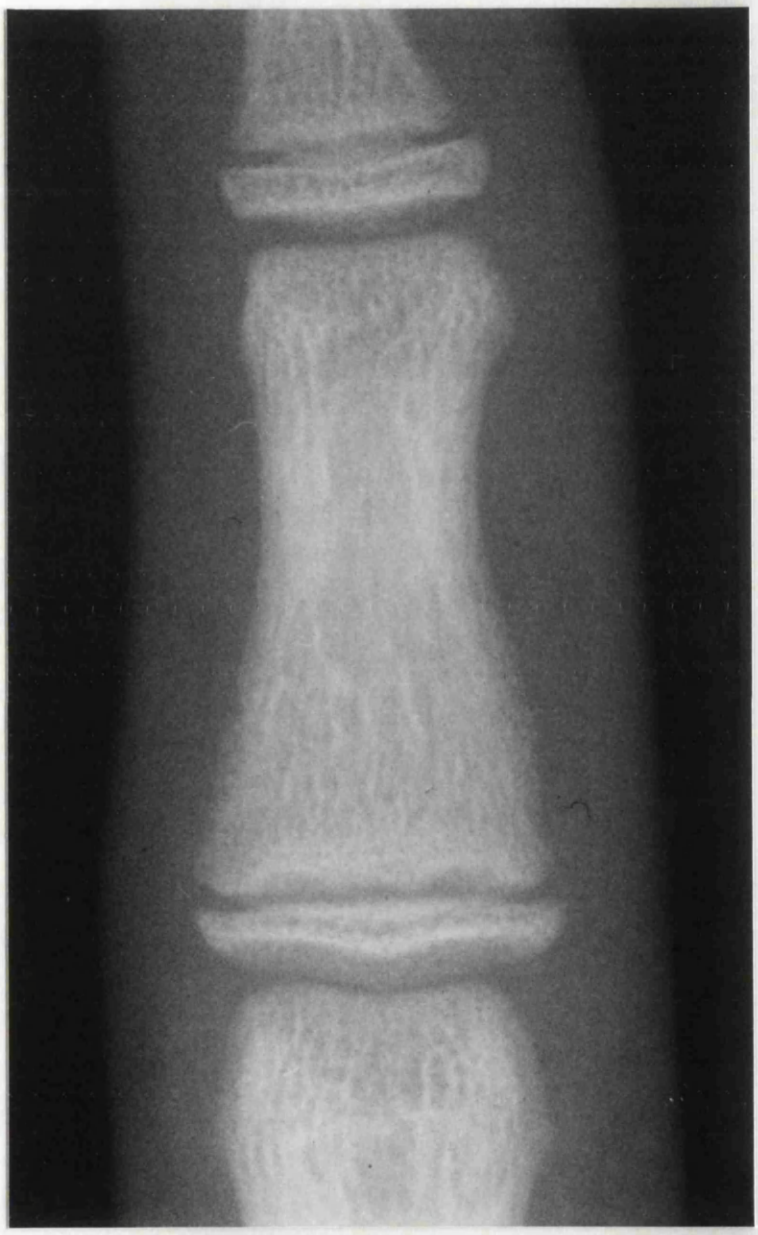
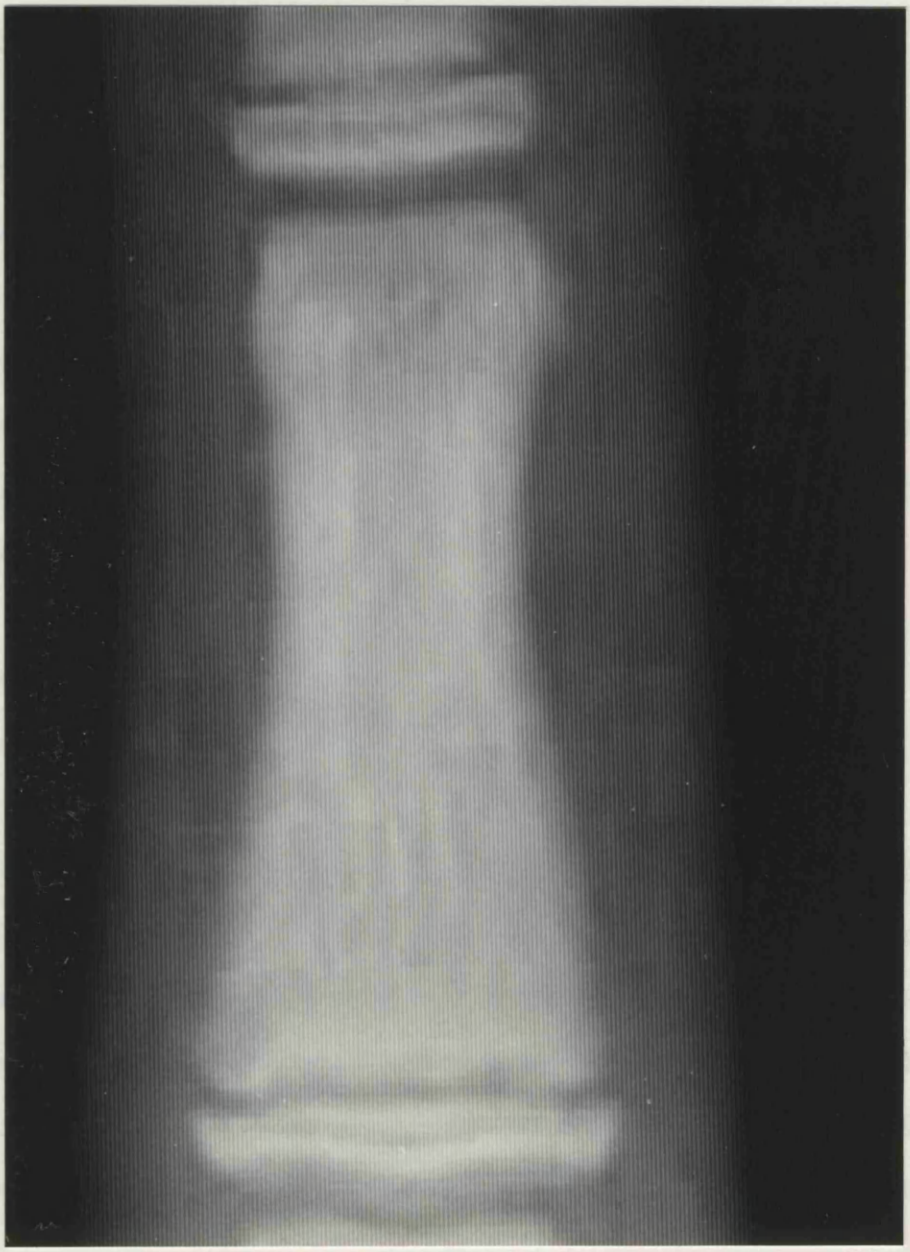


Figure 6.6: displayed images from a patient with chronic renal failure: "2K" with compression. Note blocky artifact.



6.3. SYSTEM 2

6.3.1. Sub-periosteal resorption

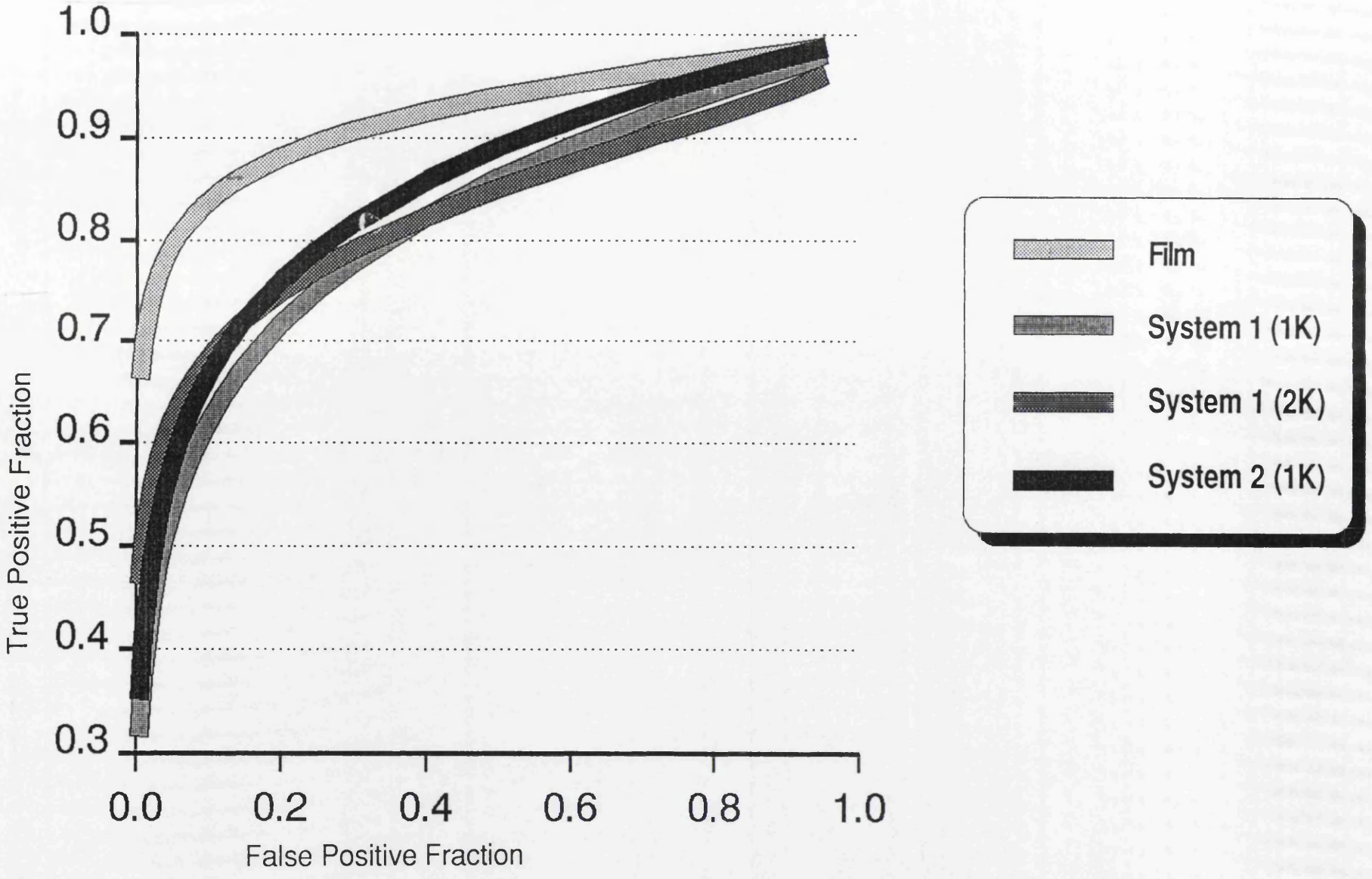
Figure 6.7 shows the ROC curves for the hand images presented on the two systems.

Table 6 (page 152) shows the true positive fractions at a false positive rate of 15%. Note that differences in figures compared with Table 5 arise from differences in the number of observers.

Table 7 (page 152) gives the significance of the differences between the curves.

There is a significant difference between film and the displayed images on both systems. The display from System 2 was better than the "2K" and "1K" formats from System 1, though these differences fail to reach statistical significance. Figure 6.5 d) (page 153) shows an image from a patient with chronic renal failure, displayed on System 2.

Figure 6.7: renal osteodystrophy images: R O C curves for system 1 and system 2.



6.3.2. *Pneumocystis carinii* pneumonia

The R O C curves obtained are shown in Figure 6.8 and Figure 6.9. Examples of the images obtained are demonstrated in Figure 6.10.

The slight difference between the curves for the short series of images that was shown twice, is not statistically significant, which again suggests that the results are consistent and reproducible.

There is a difference between the curves for film and the displayed images. The difference is significant ($p < 0.01$), but is relatively small. For example, at a false positive rate of 15%, the true positive fraction is 72% for film, and 67% for the displayed images, a difference of only 5%. The areas under the curves are 0.8650 and 0.8286 respectively.

6.3.3. PCP: with data compression

The PCP curves together represent almost 2,800 observations. The curves for the displayed compressed and non-compressed images are shown in Figure 6.11. Interestingly, the compressed images appeared to perform slightly better than the uncompressed images.

There is no statistically significant difference between the curves for film and for the displayed compressed images. Interestingly, however, the small differences between the curves for film and the uncompressed images, and between the compressed and uncompressed images, are just statistically significant ($p < 0.01$ and $p < 0.03$ respectively). At a false positive rate of 15%, the true positive fraction is 72% for film and for the compressed images, and 67% for the non-compressed images.

Figure 6.8: Pneumocystis carinii pneumonia: R O C curves for display, first and second viewing.

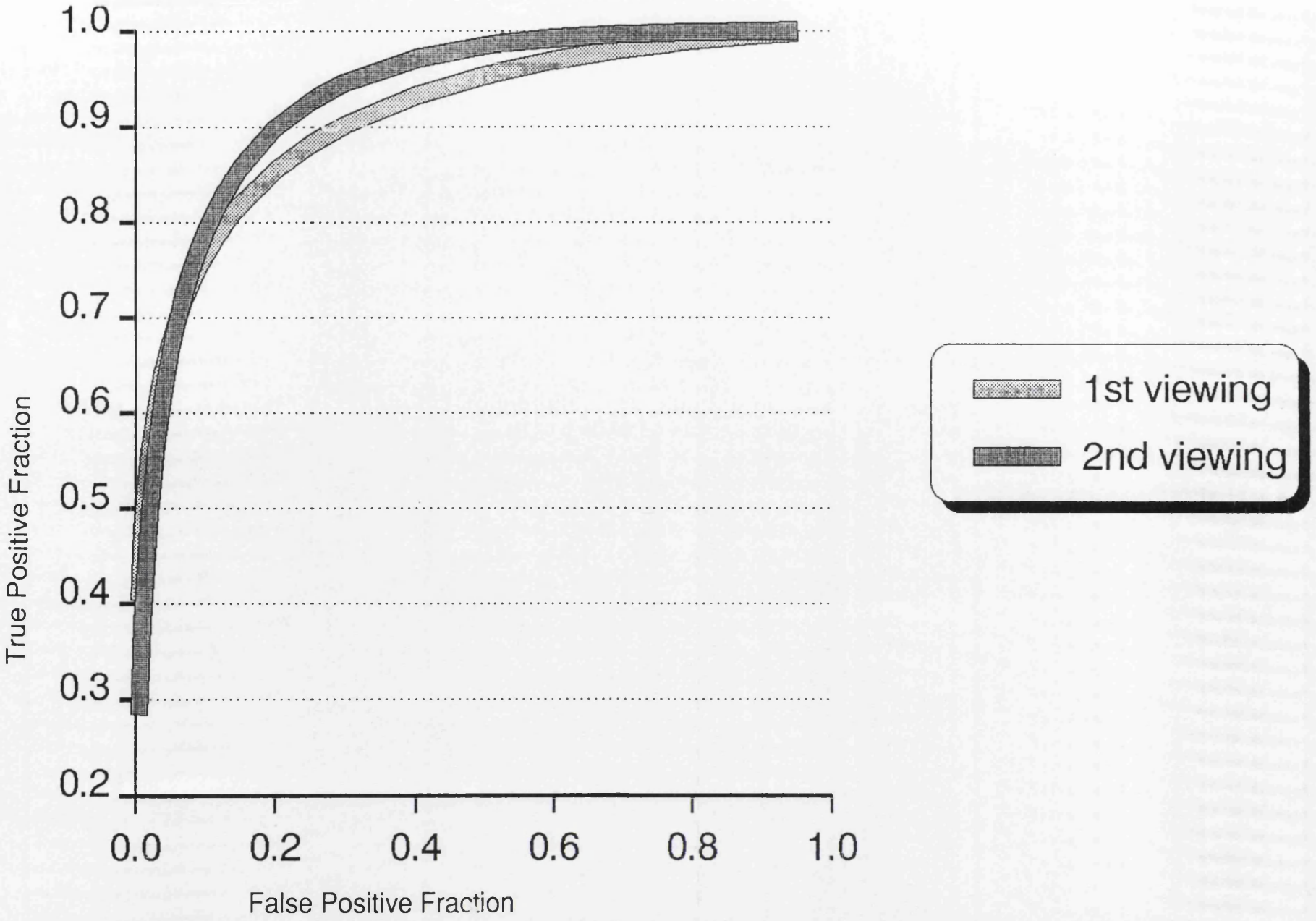


Figure 6.9: Pneumocystis carinii pneumonia: R O C curves for film and displayed images.

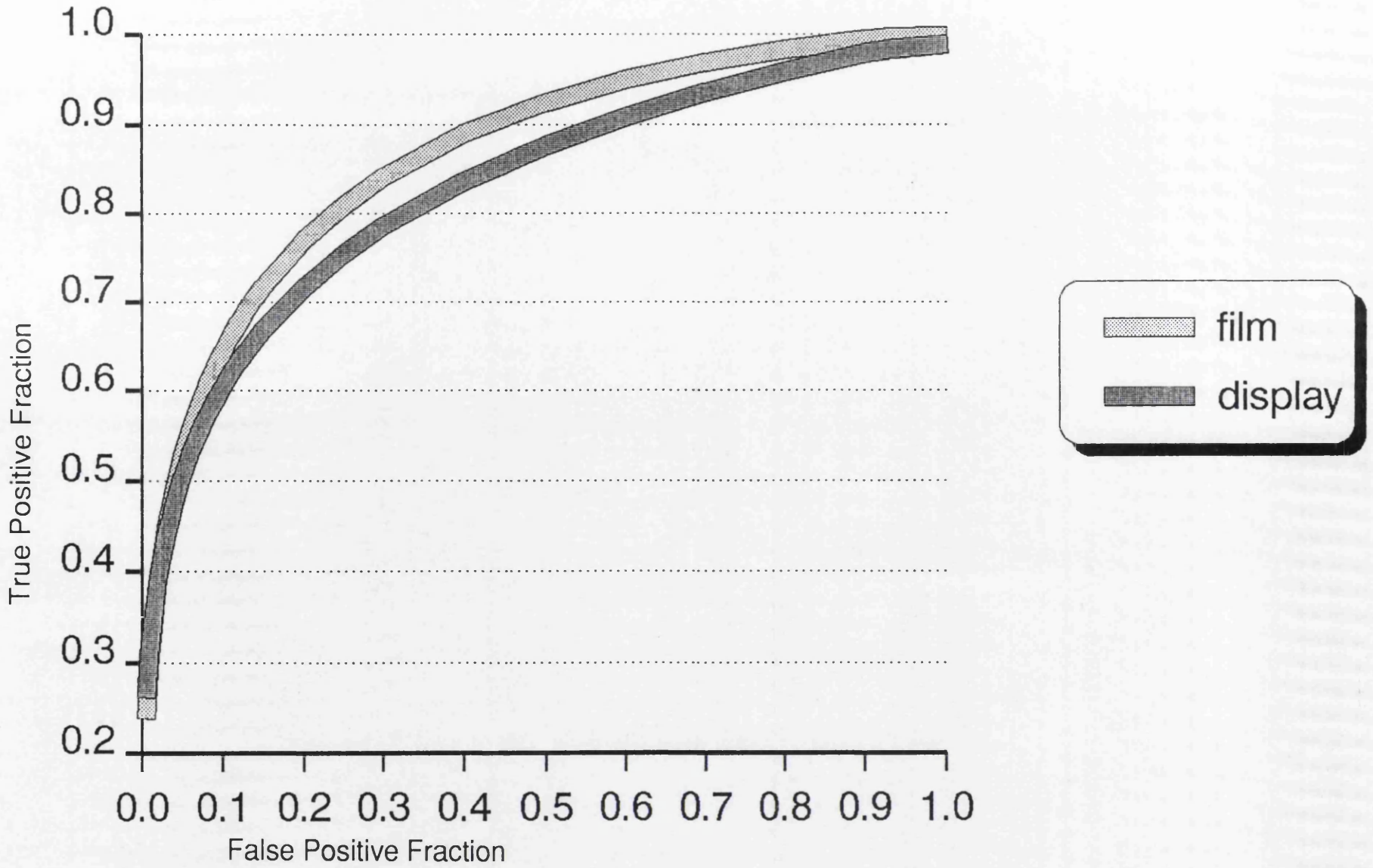
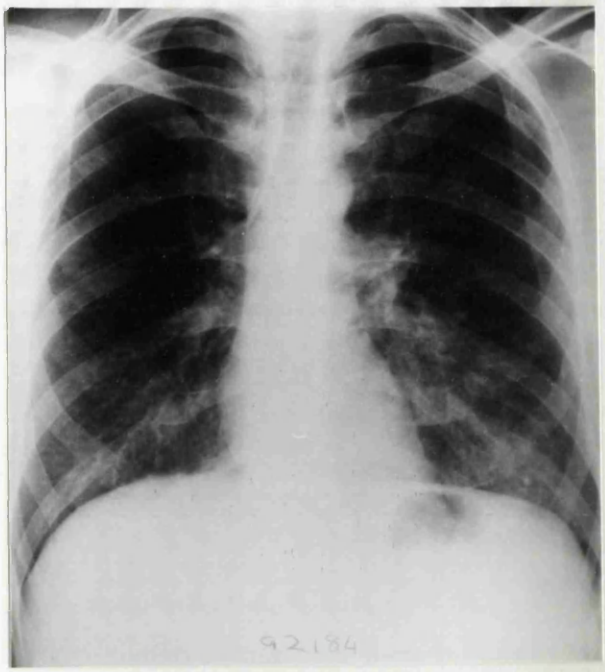


Figure 6.10: chest radiograph from patient with Pneumocystis carinii pneumonia: a) film (top); b) display.



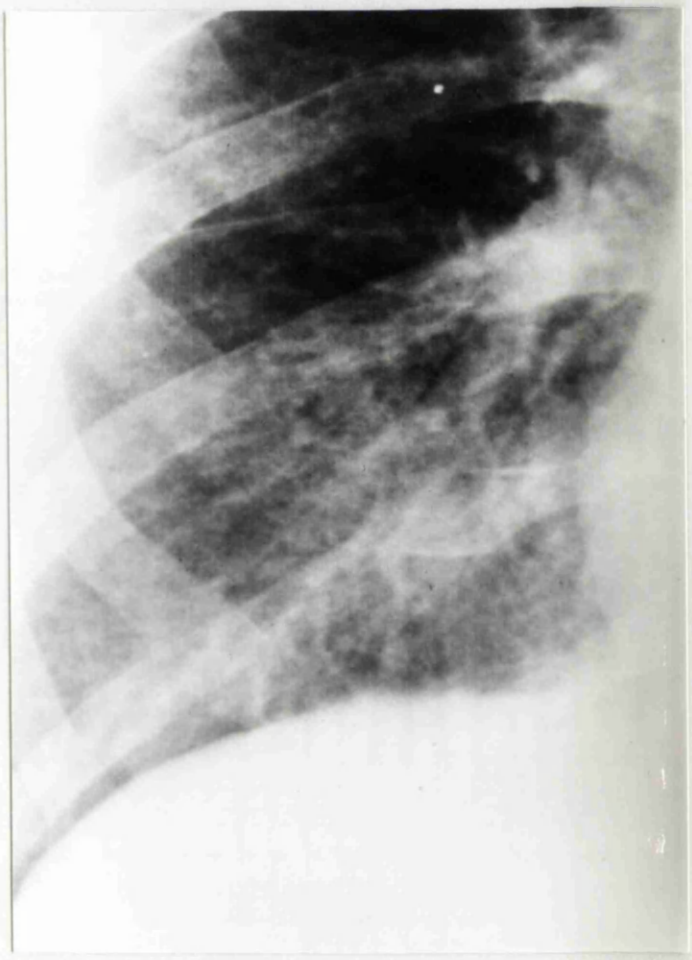
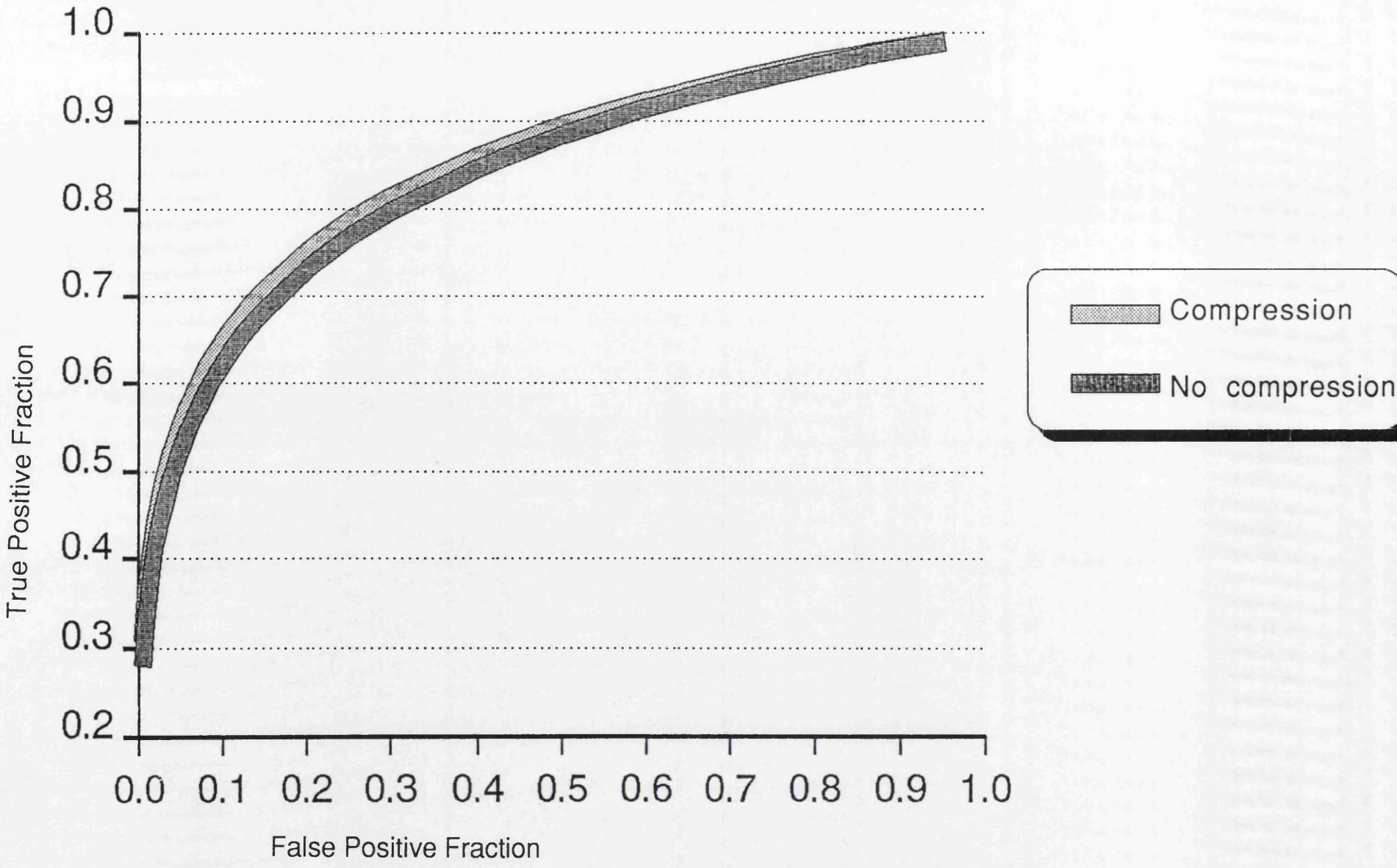


Figure 6.11: Pneumocystis carinii pneumonia: R O C curves for displayed images, with and without data compression.



6.3.4. Mammography

For the mammographic images, the resulting ROC curves are shown in Figure 6.12. The difference between the curves for film and the displayed images is substantial, and is highly significant ($p < 0.0001$). For example, at a false positive rate of 15%, the true positive fraction is 74% for film, and 54% for the non-compressed images. The areas beneath the curves are 0.8536 and 0.7200 respectively.

Figure 6.13 shows examples of the images.

6.3.5. Mammography: with data compression

The ROC curves (Figure 6.14) show no significant difference in diagnostic performance between the compressed and the uncompressed images. At a false positive rate of 15%, the true positive fraction is 52% for the compressed images - like the non-compressed images, very poor in relation to film. The areas under the curves are 0.7203 and 0.7177 respectively.

Figure 6.12: microcalcification: R O C curves for film and displayed images.

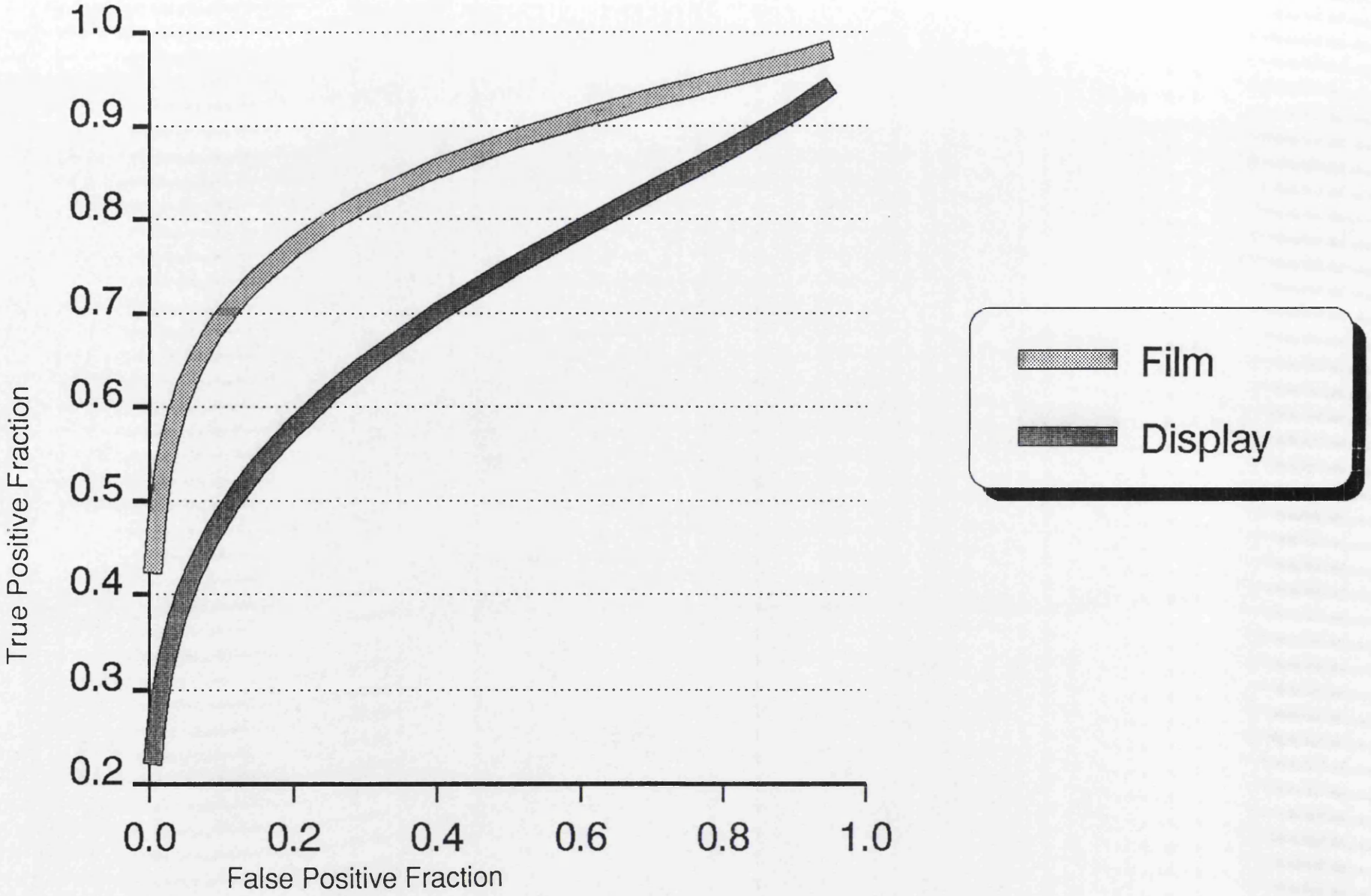
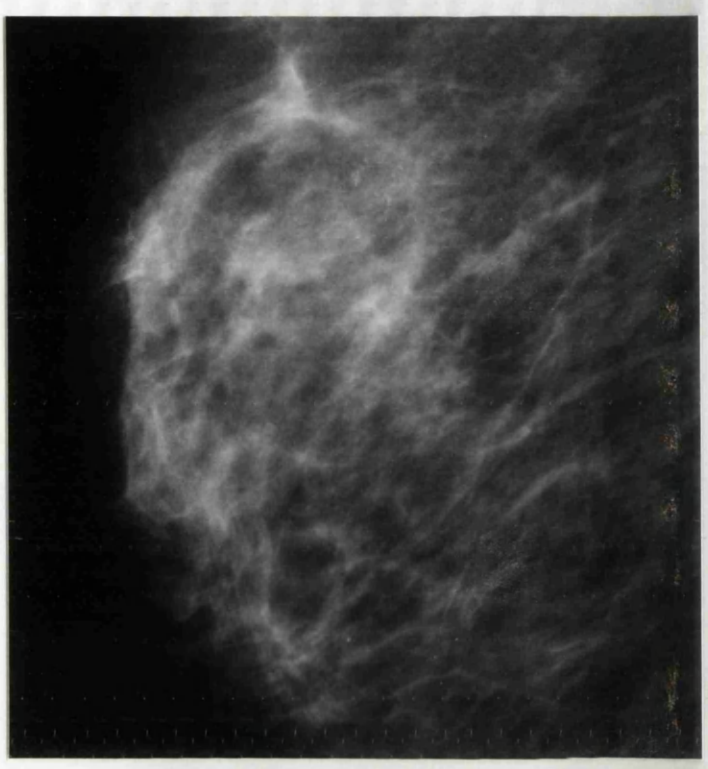


Figure 6.13: microcalcification: a) film and b) displayed images.

a



b

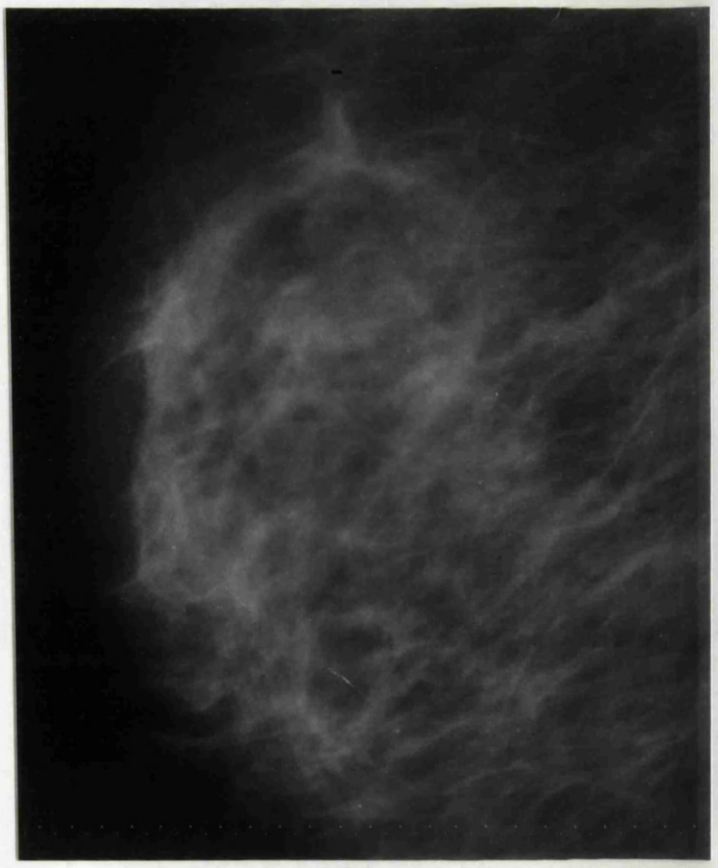
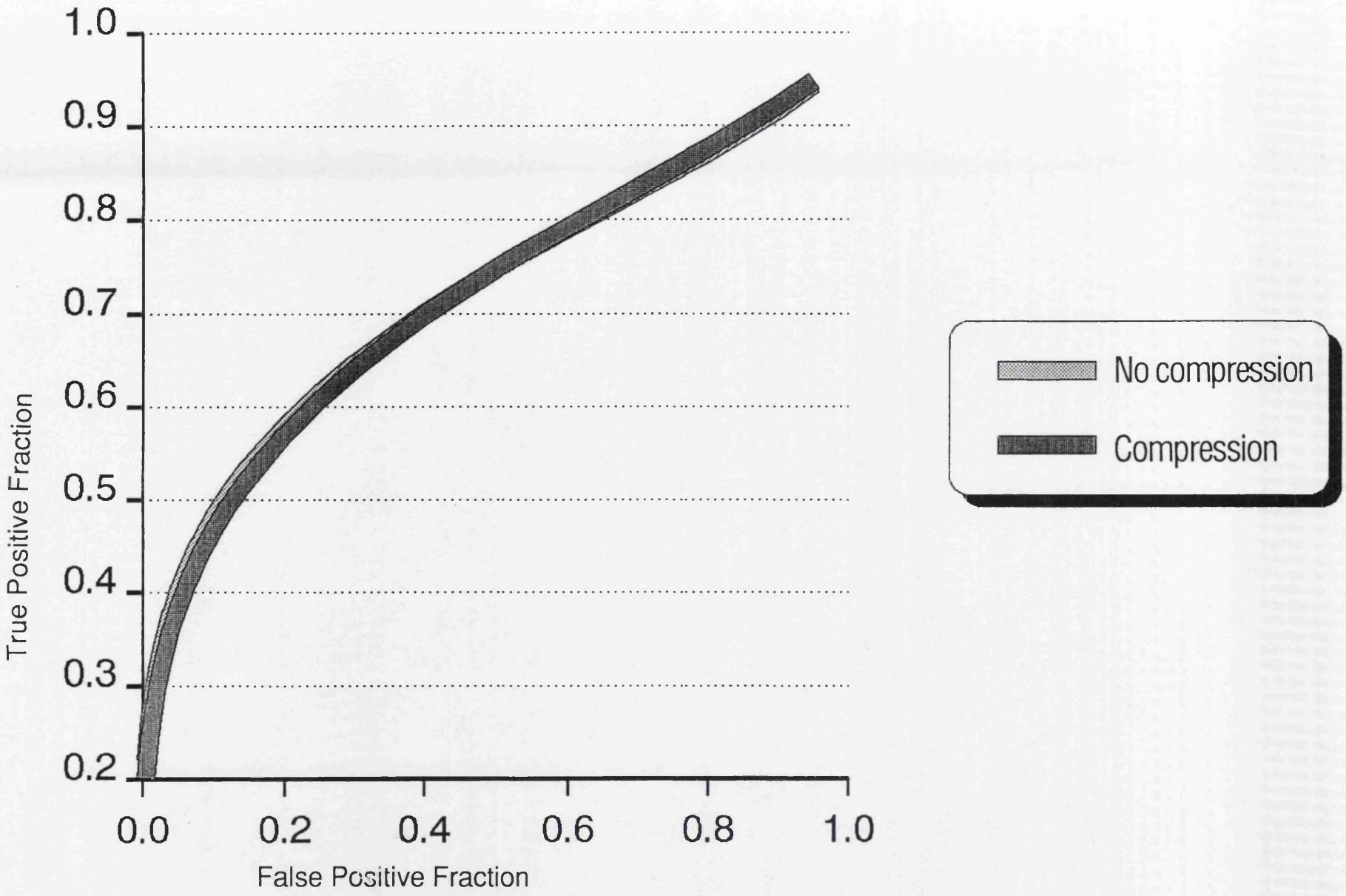


Figure 6.14: microcalcification: R O C curves for displayed images, with and without data compression.



6.3.6. Skull fractures

A set of ROC curves was first plotted using the assumption that **all** of the images from the patients with skull fractures were abnormal; in other words, if a fracture was clearly visible on a lateral view, the frontal views were also assumed to be abnormal even if the fracture was not readily visible (Figure 6.15). This assumption was clearly unreasonable, since some of the skull views of these patients actually omitted the region of the fracture from the film. Not surprisingly, the effect of this assumption was to introduce a random element into the responses, thereby flattening both the curve for film and the curve for the displayed images. The difference between the curves was nonetheless statistically significant. (Areas under the curves were 0.6730 and 0.6346 respectively.)

A second approach would have been to seek a "panel" verdict, not just for each patient, but also independently for each film, such that films on abnormal patients that failed to show the fracture would be excluded. Data was collected with such an exercise in mind, but this approach was ultimately considered to be too subjective to be worthwhile.

More important than the normality or abnormality of each image is the conclusion that can be drawn from the entire image set for each patient. A computer programme was therefore written (Andrew Todd-Pokropek) to derive from the

data the most abnormal response for each patient from each observer. The ROC curves derived from the pooled data are shown in Figure 6.16.

The difference between the curves for film and for the displayed images is again statistically significant ($p<0.005$). For a false positive fraction of 15 per cent, the correct diagnosis would be made in 67 patients using film, but only 59 patients using the displayed images. The areas under the curves are 0.8321 and 0.7770 respectively.

There was no significant difference between the curves for the short sequence of displayed images shown twice.

Figure 6.15: skull fractures: R O C curves for film and displayed images (with **all** images from patients with fractures assumed to be abnormal.)

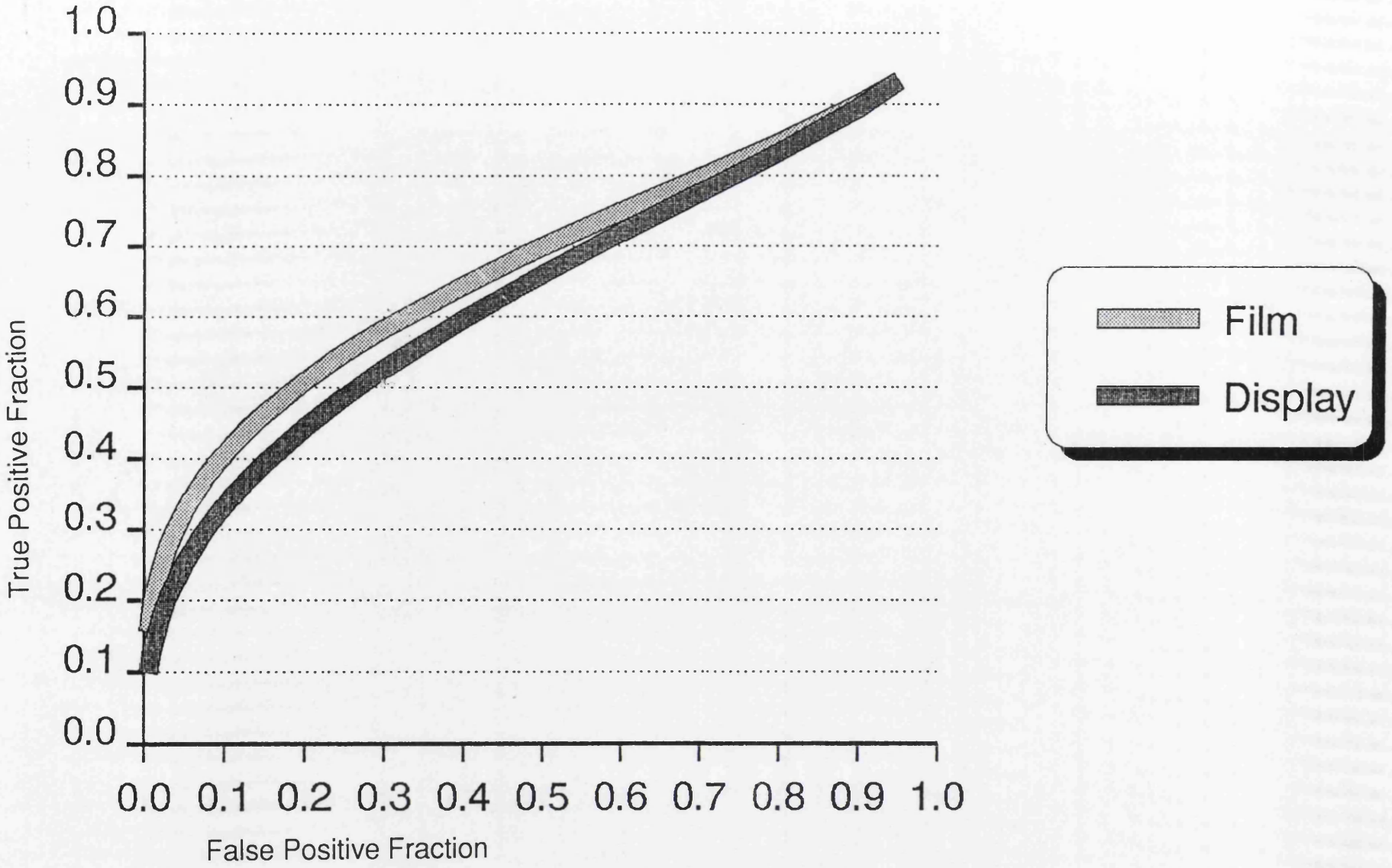


Figure 6.16: skull fractures: R O C curves for film and displayed images (based on each observer's most abnormal response for each patient).

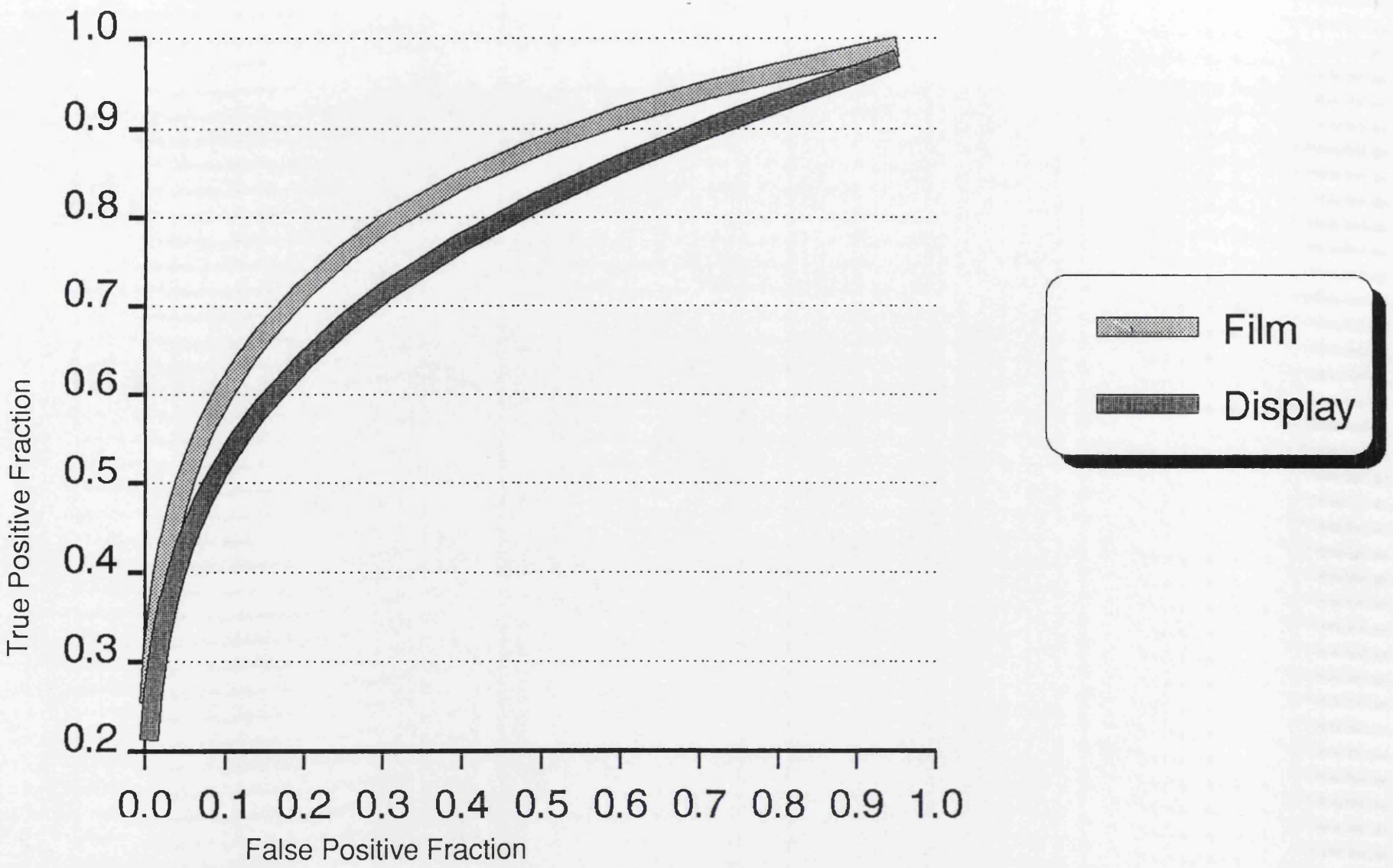
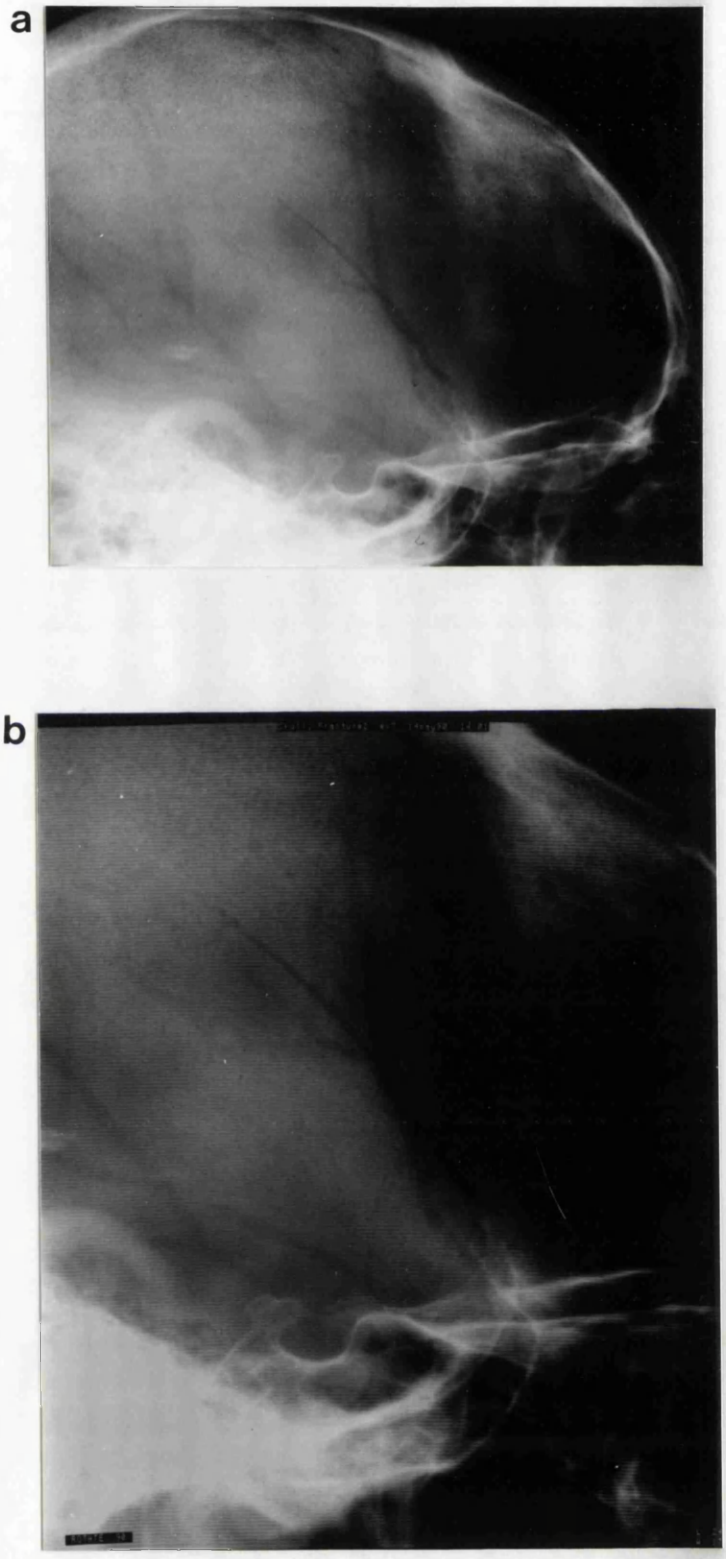


Figure 6.17: Skull fractures:
a) film and b) displayed images.



6.4. VIEWING TIMES

Our data capture software automatically recorded time taken to view each image. The times taken to complete 6 representative viewing sessions - chosen from a period when all observers had been using the equipment for several months (System 2) - are presented in Table 8 on the next page: times taken to complete identical observation tasks with the display system are typically three to four-fold longer than with film.

The major factor in these delays relate to the speed of image retrieval: the systems studied were slow, and resulted in frustrating delays between images. Manipulation of the images - particularly and scrolling around zoomed images, and alteration of window settings, introduced a further interpretation delay.

TABLE 8: Times taken per observer per viewing session (minutes): PCP and mammograms.

	PCP			Breast		
	film	-comp	+comp	film	-comp	+comp
Observers 1:	30	127	100	21	76	82
2:	27	116	74	26	90	76
3:	38	192	173	23	129	87
4:	43	140	122	49	97	88
5:	24	127	84	-	-	-
6:	32	139	103	39	141	104
7:	21	151	78	22	82	61
8:	17	116	82	-	-	-
9:	30	147	150	28	120	118
10:	-	-	-	30	68	77

PCP = Pneumocystis carinii pneumonia on chest images (113 cases); Breast = mammographic images for microcalcification (80 cases); +comp/-comp = displayed images, with or without data compression.

7. PHYSICAL EXPERIMENTS

7.1. INTRODUCTION

At the start of this evaluation project, the principal purpose had been to examine the effects of matrix size and data compression; it quickly became clear, however, that there were other factors at play in determining displayed image quality. It became important to find ways of examining their influence and significance.

Tests were performed to explore some of these physical characteristics, and also the quality of data acquisition by film digitization, using phantoms and test objects.

7.2. FLICKER AND JITTER

Flicker and jitter are particularly disturbing, especially to peripheral vision: these were studied first.

The extent of perceptible flicker with System 1 had been a source of concern since first installation; its impending replacement with System 2 afforded the opportunity to make some simple comparative measurements.

The term flicker refers to those temporal variations in grey level output perceived by the eye at frequencies of 1-

100Hz. The term jitter is used to represent spatial motion of the displayed image at similar frequencies.

It was considered that the essential measurement to be made was that of light output from the phosphor of the display. The basic method used was to place a fast photodiode (BPX65, Radiospares Components) looking at light output over small regions. This diode was placed at a fixed distance from the screen; it incorporated a lens in front of the active surface, limiting the area from which light could be detected to a region of about 2mm x 2mm. A conventional 100Mhz oscilloscope was used to determine high frequency effects (such as phosphor decay) and a frequency analyzer was used to look at low frequency effects in the region of 5Hz up to 1KHz. More sophisticated devices have been described by other authors (e.g. Roehrig et al 1989). It is also possible to look directly at the video signal being input to the monitor, and to measure such parameters as frequency response, electronic dynamic range, etc., though such experiments are not reported here.

Jitter could be determined in part from triggered information observed on the oscilloscope trace, and in part by visual estimation of displacement, looking through a small slit. In the first case, given a high contrast spot, such as part of a ASCII character on the screen, it was possible to trigger the oscilloscope to the signal from a single video pixel. Frame by frame variations could then be observed. In the second case, by use of the slit and a

suitable high contrast detail, the physical shift of the position of the detail across (or down) the screen could be assessed.

Flicker, however, is more difficult to measure. What appears to be most disturbing is variations in intensity at relatively low frequencies in the region of 1Hz up to 100Hz. These were assessed by looking at the frequency power spectrum. Various other features were observed, in particular, the refresh rate, the interlace (when present) and also interference from 50Hz (mains) frequency which seems to be impossible to eliminate - especially in proximity to other electrical equipment. For comparison, three other monitors from different manufacturers were also tested, installed at different sites, the first being a 512 x 512 display running at 60Hz, the second being 800 x 600 at 50Hz, and the final system being a 640 x 480 display, running at 50Hz. (Note that both System 1 and System 2 were running at 60Hz, the standard in the USA, rather than the normal 50Hz European standard.)

Grey level uniformity was measured by looking at the output from the diode at various positions, and also monitored by using a (slow) light cell based densitometer, also moved manually. An alternative would be to look at the display with an appropriate TV camera system, after corrections for non-uniformity of the camera itself (Roehrig et al 1989).

Resolution along the video line can be measured by looking at the oscilloscope trace from a single bright dot on the monitor. Resolution is primarily determined by the electronic performance of the display and the phosphor temporal and spatial characteristics of the display screen.

Most such measurements do not require specialized equipment, and can be performed with a simple microcomputer and analogue to digital converter (ADC) connected to the photodiode.

Figure 7.1 shows the frequency spectrum observed from the System 1 display. Note the 50Hz (mains frequency) peak which is always present, and (in part) comes from pickup and from the instrumentation itself. The 30Hz peak comes from the interlacing of the display. Note the relatively broad spread of the 60Hz peak.

Figure 7.2 shows the corresponding frequency spectrum obtained from the 60Hz non-interlaced display from System 2. The 30Hz peak has disappeared, and the 60Hz peak is much sharper. The broadening of the 60Hz peak was associated with low frequency components in the 1-5Hz range, precisely the low frequency flicker and jitter that was disturbing to the eye, and that we were attempting to measure.

Figure 7.3 shows these measurements in progress, with the diode against screen.

Figure 7.1: frequency spectrum from System 1 display. Note the 50Hz mains frequency peak, the 30Hz peak from the interlacing of the display, and the broad spread of the 60Hz peak.

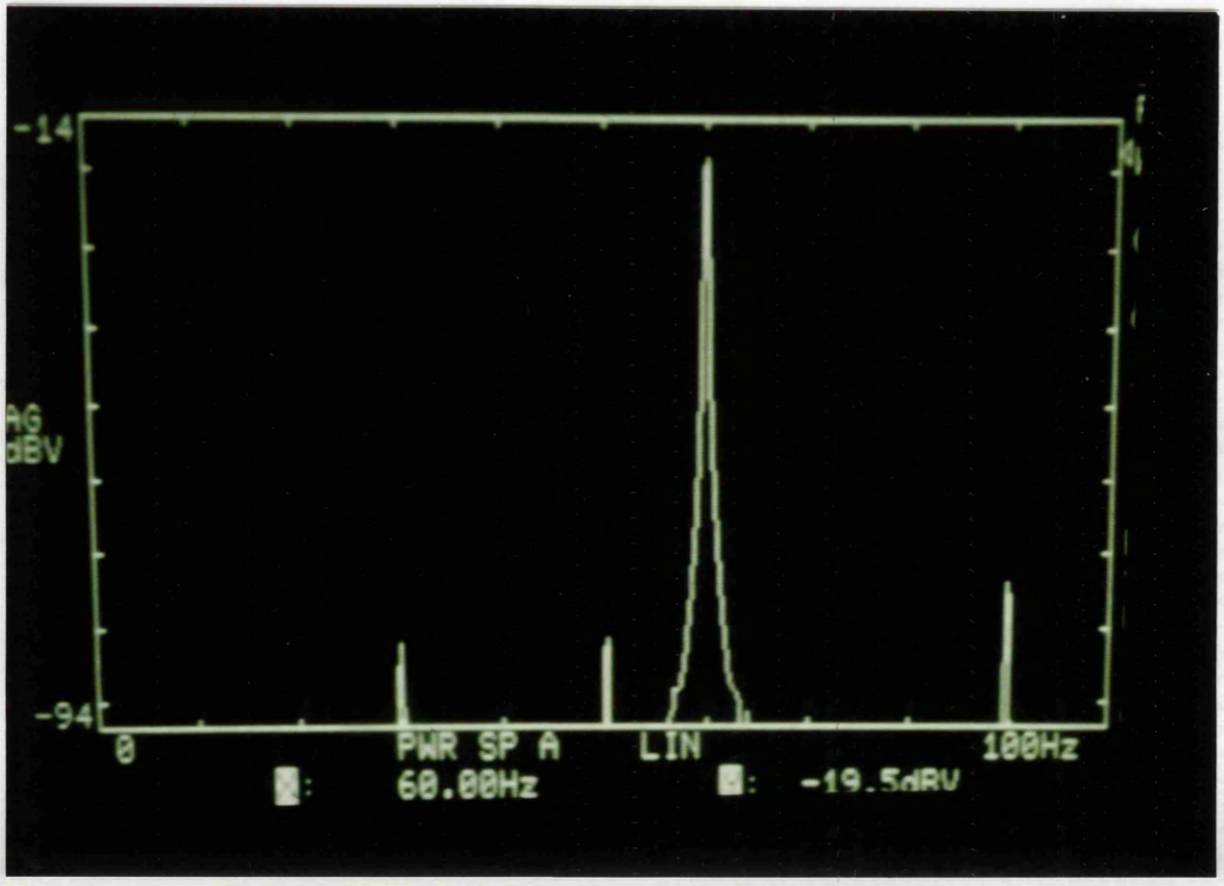


Figure 7.2: frequency spectrum obtained from System 2 display. Note that the 30Hz peak has disappeared, and the 60Hz peak is much sharper.

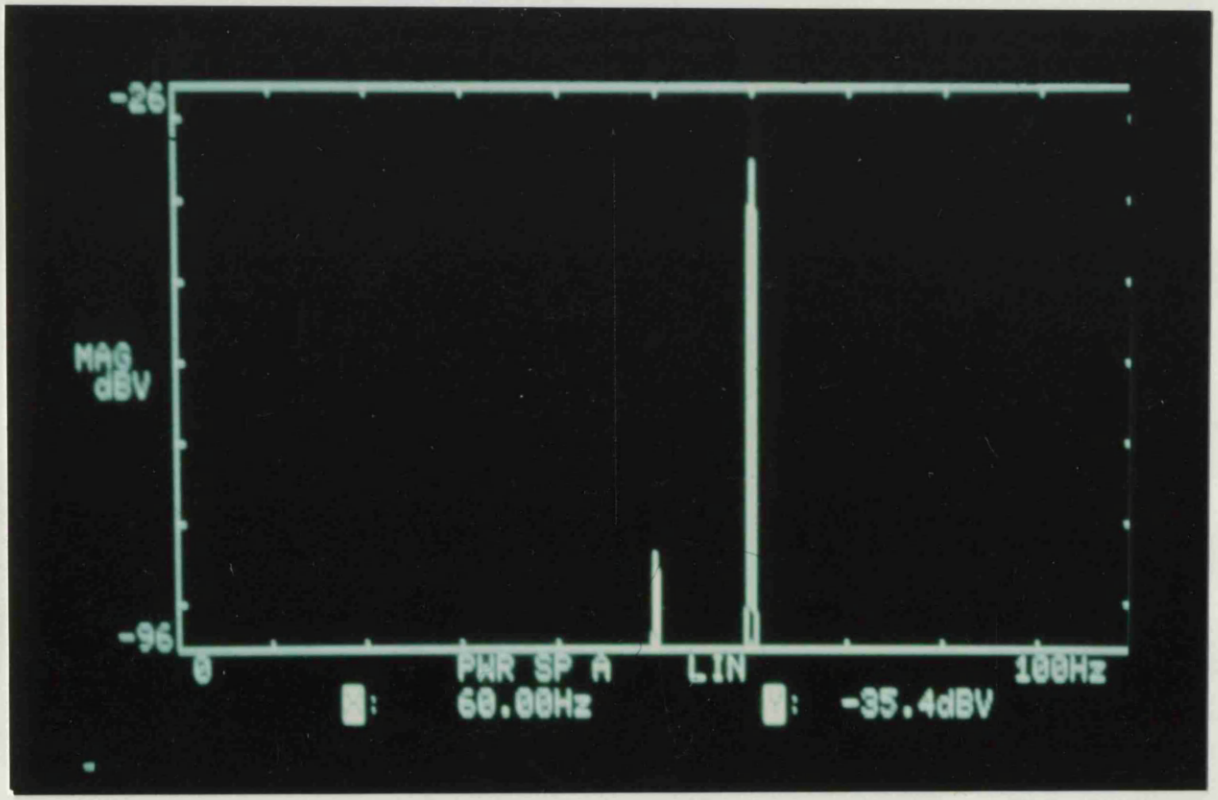
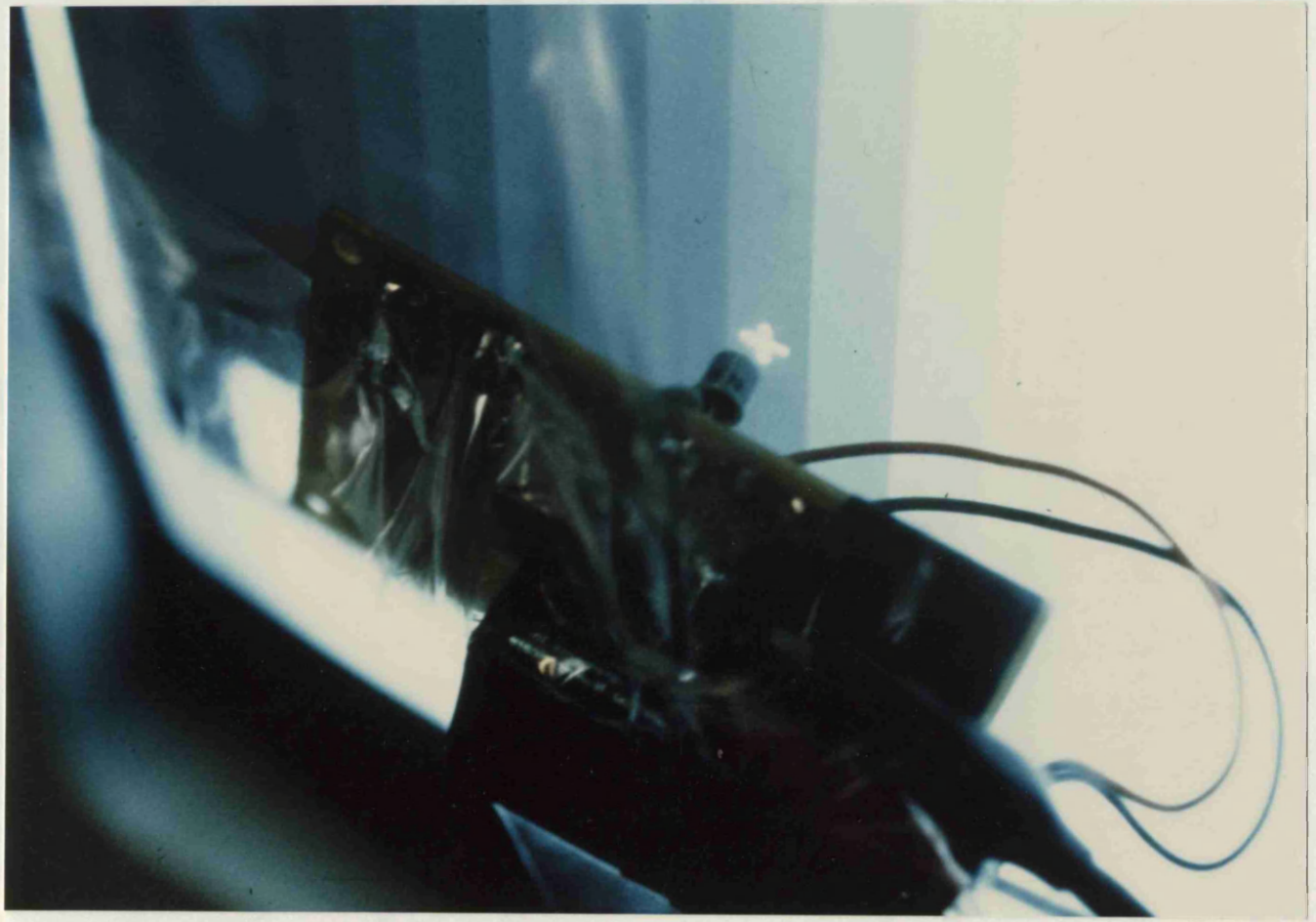


Figure 7.3: measurements in progress, with diode against display screen.



Resolution can be determined by using a photo-diode to look at the fast response. Alternatively, synthetic test patterns can be inspected by an observer for perceived resolution - and these ought to be standard issue from manufacturers.

Grey level uniformity, signal to noise ratio, distortion, changes in aspect ratio, detail in the display, artifacts, video line cross section, and phosphor decay time, are also worthy of measurement, as well as the colour temperature of the display (colour temperature can be defined as the absolute temperature at which a black body radiator would have a chromaticity equal to that of the light source) and its uniformity.

7.3. CONTRAST RESOLUTION

The contrast resolution properties of the system were investigated using a home-made low contrast phantom. This consisted of a number of sheets of blank, conventional radiographic film, with holes of different sizes cut from different numbers of sheets which were then superimposed (Figure 7.4). The sizes of the "lesions" ranged for 1cm X 1cm down to 1mm x 1mm. This phantom was radiographed through a variable amount of scattering material, to create a set of 40 low contrast images.

The resulting radiographs, one of which is shown in Figure 7.5, were then digitized (System 2); the images were

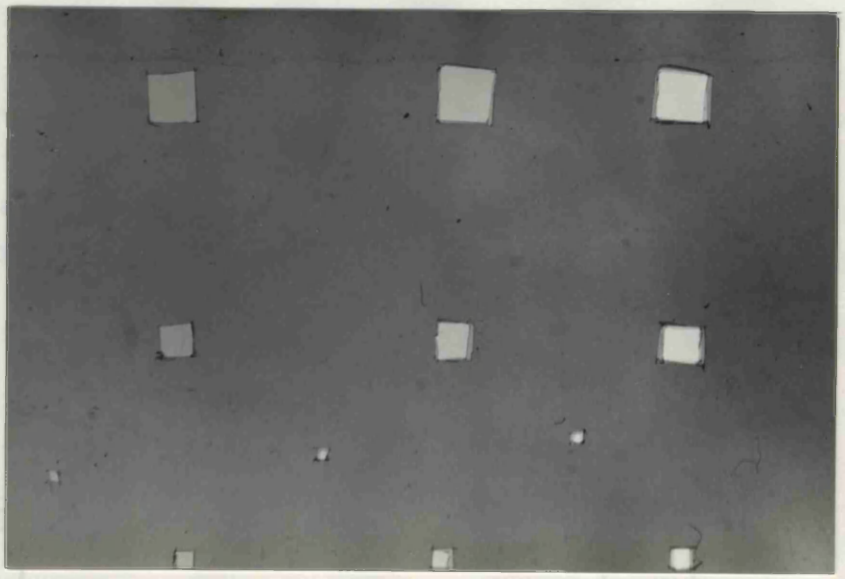
presented to observers who were requested to indicate how many lesions they could see on each image, both on film and on the digital display. For each pair of images (digitized and on film) the difference between the number of lesions visible was recorded and used in statistical analysis. The images were interpreted by a total of 6 individuals.

It is important to note that, on the displayed digitized images, which were essentially uniform, the observers were allowed to use the window functions of the display to enhance the contrast of the image. The main limitation to such contrast enhancement was a residual non-uniformity of the digitized film, and the noise of the digitization process. These lesions were in no way limited by spatial resolution. Thus the main purpose of this experiment was to test the noise properties of the digitizer and the perception of contrast in a noisy image.

The results, which are summarized in Figure 7.6, show that lesion detectability on the monitor was significantly better than on the original film; but they did not appear to be critically dependent on the window levels that the observer chose; this is in keeping with the findings of others (for example Judy et al 1989).

In Figure 7.6, the histogram plots the number of additional lesions seen either on film or on the digital display. This histogram is (significantly) skew towards increased lesion detection on the digital display.

Figure 7.4: photograph of low contrast phantom.



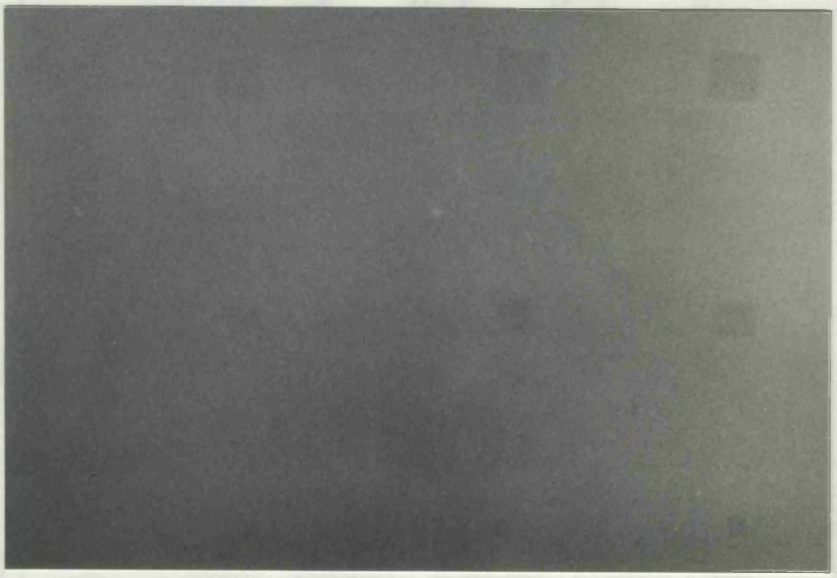
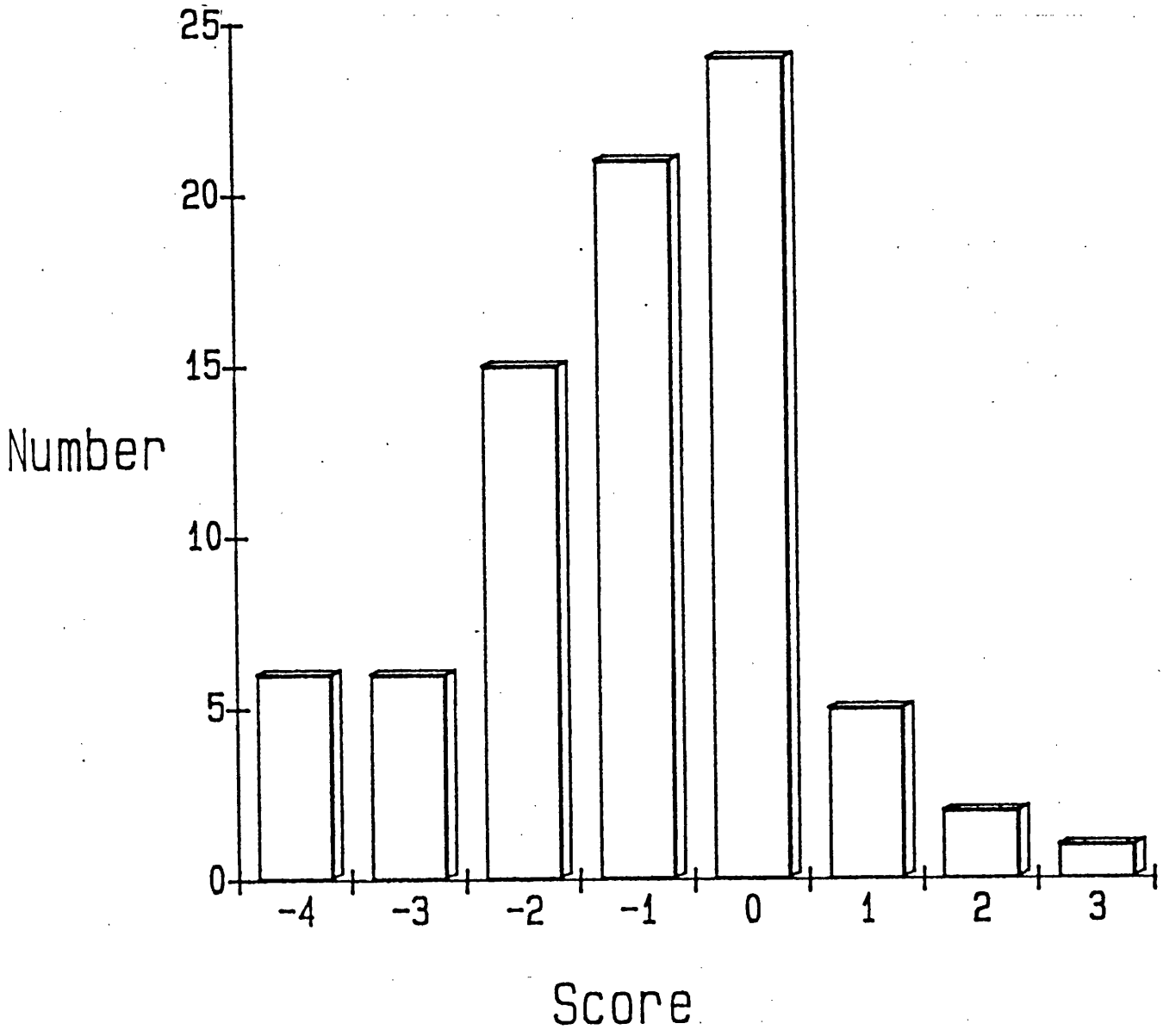


Figure 7.5: print of radiograph of low contrast phantom.

Figure 7.6: histogram of low contrast lesion observations. The histogram shows the number of images in which a different number of lesions was detected between film and the digital display. A score of zero indicates concordance, a negative score indicates that more lesions were observed on the digital display, and a positive score indicates that more lesions were observed on film. The difference between film and digital display is highly significant.

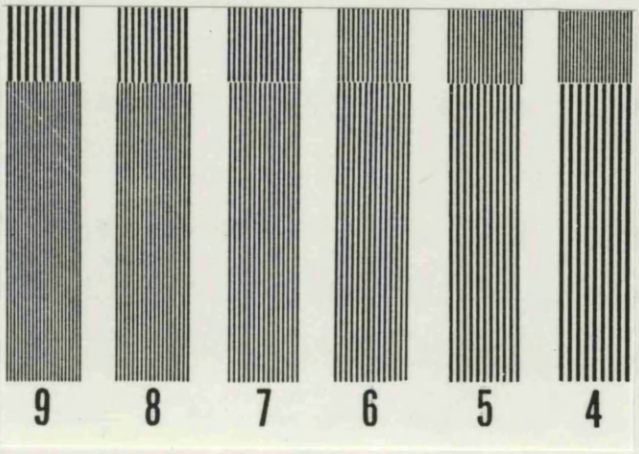


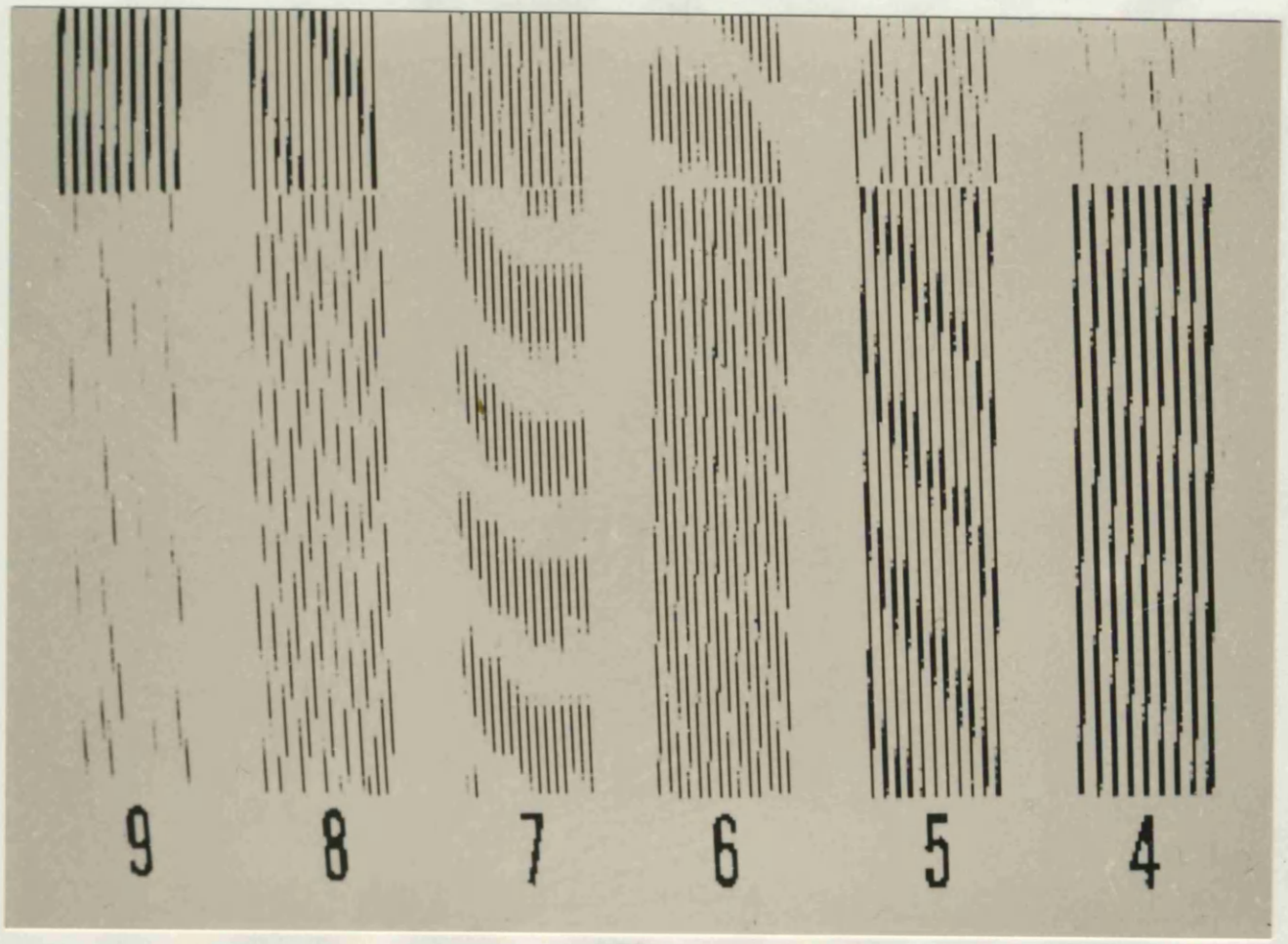
7.4. SPATIAL RESOLUTION

It was clear that current PACS display systems have a spatial resolution that is markedly inferior to that of conventional film, and we attempted to investigate this with System 2.

A digitizer scanning spot size of 210μ ought to yield a spatial resolution of approximately 2.4 line pairs/mm. Digitized test patterns enabled clear discrimination of just 0.9 - 1.1 line pairs/mm (Figure 7.7) - though moiré patterns complicate such measurements.

Figure 7.7: spatial resolution test pattern: a) original; b) digitized display (below). Line pattern "4" corresponds to 0.9 line pairs per mm; "5" corresponds to 1.1 lp/mm; "6" corresponds to 1.4 lp/mm; "7" corresponds to 1.5 lp/mm; "8" corresponds to 1.8 lp/mm; and "9" corresponds to 2.1 lp/mm.





7.5. SIGNAL TO NOISE RATIO OF DIGITIZER OUTPUT

This experiment was concerned with the influence of a laser film digitizer on the system, and the number of effective grey levels that could be created (i.e. the dynamic range of the system).

The signal to noise ratio and the number of useful bits in data derived from laser film digitization was assessed for different film densities, using System 2. A set of six test films of different optical density ranging from 0.8 up to 2.5 was digitized; in addition a blank (filmless) image was also digitized. It was ensured that the test films were uniform and noise-free.

Initial tests had been performed using X-ray film exposed to give different densities, but these images were found to be very noisy and therefore quite unsuitable for testing the digitizer. Much of this noise was attributed to film processing, and to non-uniformity of the radiographic beam. In order to avoid these problems the test images used here were generated from different thickness of **unexposed** film.

After digitization, the raw image data was downloaded to magnetic tape, and transferred to an external computer; this task was made needlessly complex by the manufacturer's view that we were entering commercially sensitive territory, and by their consequent unwillingness to provide information regarding data format. (This is apparently a

widespread problem.) Mean pixel values and standard deviations within various sizes of regions of interest were computed.

7.5.1. Results

Figure 7.8 shows a graph of mean pixel value and the corresponding standard deviation for a central 256 pixel region of interest, plotted against optical density of film for the uniform test film set. The standard deviation increases as the film density increases (and as the mean pixel value decreases).

From these values, an estimate of dynamic range can be obtained, as plotted in Figure 7.9. The dynamic range was computed using:

$$\frac{(Max-Mean)}{S.D.}$$

where "Max" represents the maximum pixel value at the highest optical density, "Mean" represents the mean pixel value at the optical density considered, and "S.D." represents the standard deviation at that optical density.

The least significant bits are primarily noise. However, it has been suggested that it is desirable for noise to be perceptible in the image for good performance in detecting subtle features (Judy et al, 1981). This relates to the choice of the display window selected. Where the noise is

not perceptible, either false positive rates (as analyzed using ROC methods) tend to increase, or true positive rates to fall. Thus an increase in the dynamic range of the digitizer, defined in terms of signal to noise ratio, which would result from a decrease in digitizer noise would also result in a need for a greater dynamic range in the display, and more careful use of windowing.

7.5.2. Discussion

It may be observed from Figure 7.9 that the effective number of grey levels (the dynamic range) is less than 256 for most optical densities. Thus, although the data was supposedly being digitized to 12 bits, the least significant 4 bits contain mostly noise, and in effect only 8 bits remain. This was for a 210μm scanning spot, and the noise level would be expected to rise with a decreased spot size. The signal to noise ratio is particularly poor for high optical density, when little light is transmitted. The signal to noise ratio is a function of the amount of light recorded, which is determined by the optical density, the spot size and the sampling time.

On the other hand, experiment 7.2. demonstrated that contrast resolution with such a system is quite good. While some of the objects to be detected had much greater contrast than the noise level, the most difficult were concealed within the noise (as indicated by the estimates of noise presented above). Separate experiments have

indicated that noise needs to be perceptible to the observer for reliable detection of such low contrast lesions. The perception of noise should be a weak function of the window setting chosen (Judy et al, 1981).

Figure 7.8: mean pixel value and the corresponding standard deviation, plotted against optical density.

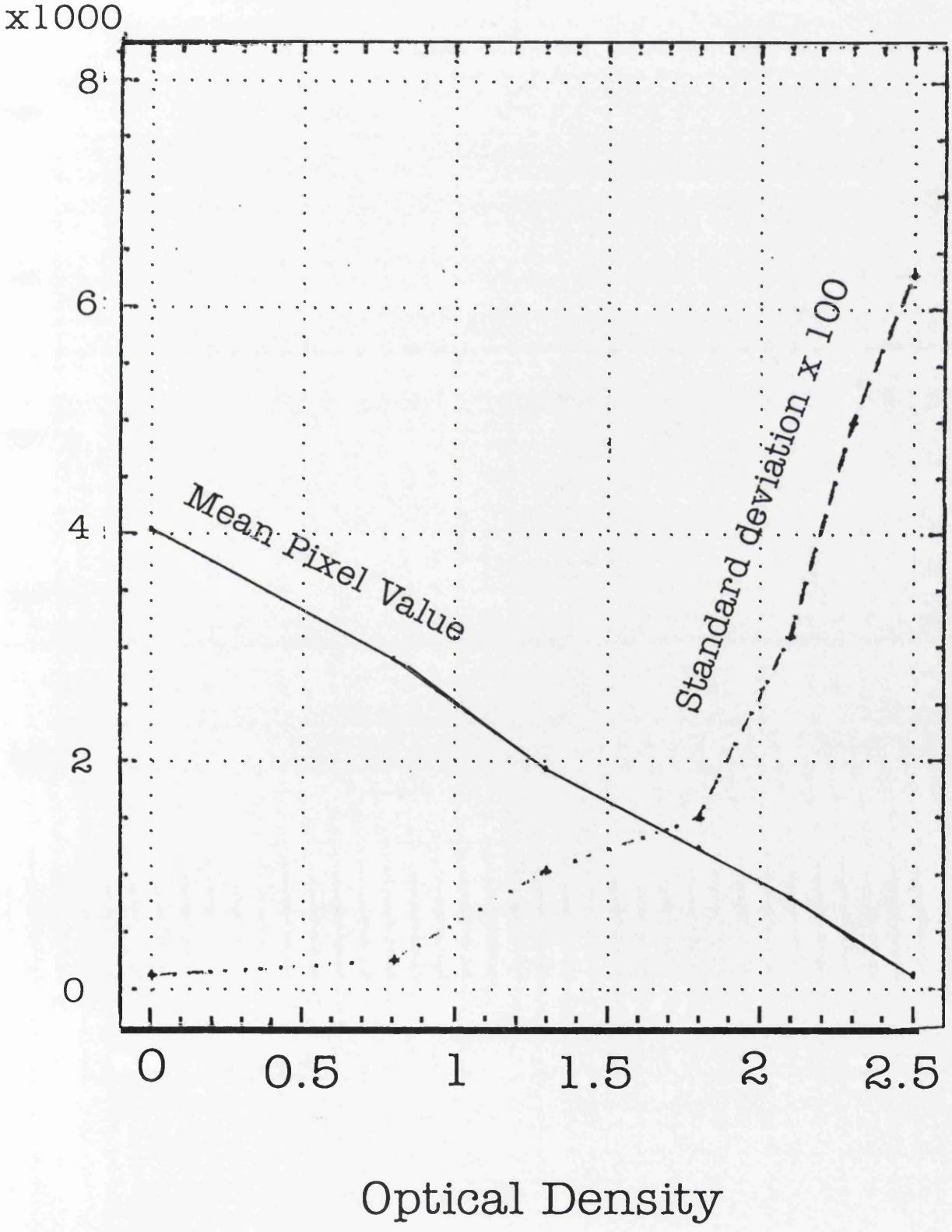
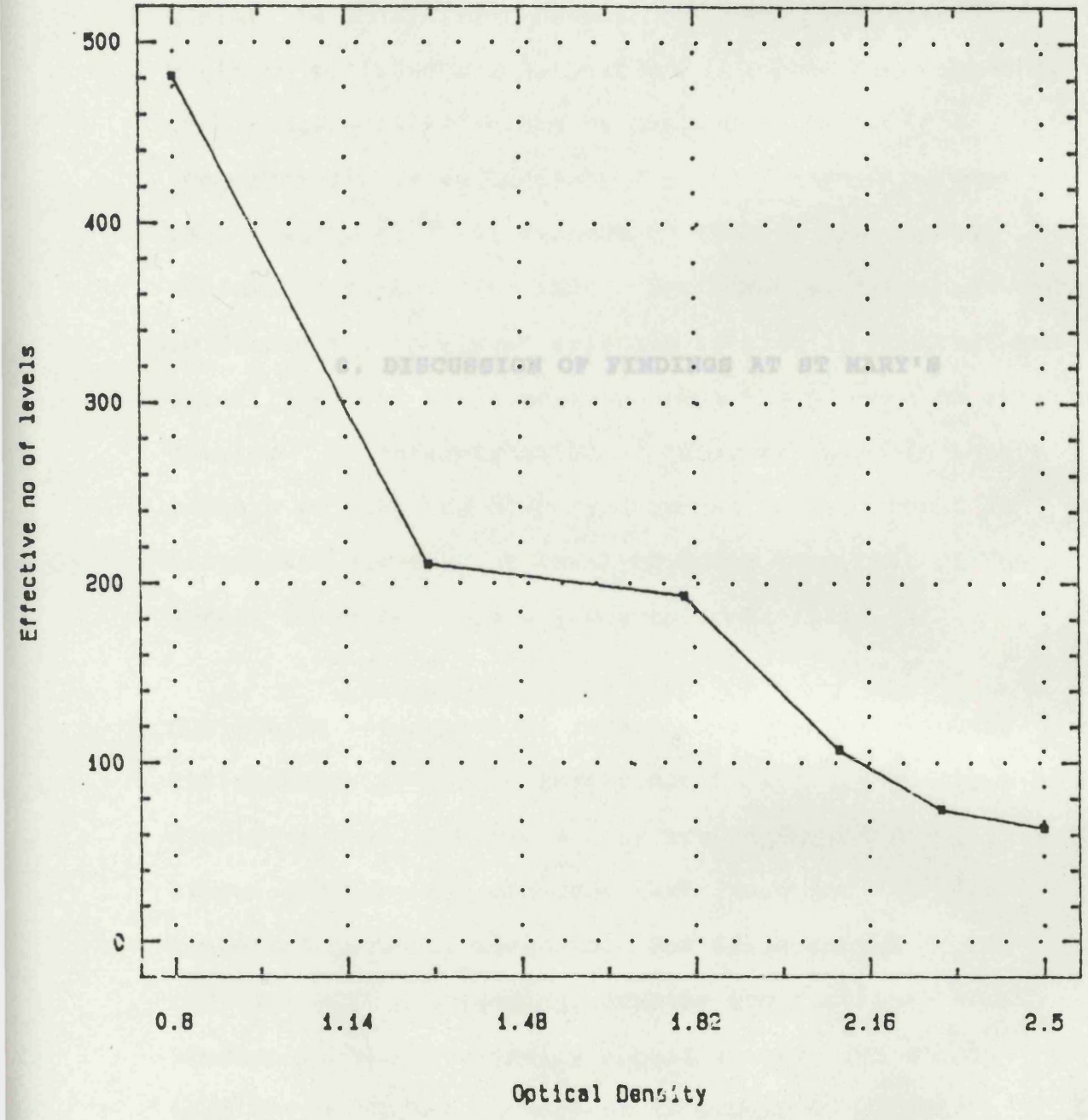


Figure 7.9: dynamic range plotted against optical density. Note that the effective number of grey levels (the dynamic range) is less than 256 for most optical densities.



8. DISCUSSION OF FINDINGS AT ST MARY'S

8.1. INTRODUCTION

There is clearly no point investing large amounts of money and effort in installing a PACS network if the displayed image quality would be clinically unacceptable. Image quality is an abstract concept for which there is no entirely satisfactory definition. Kundel has suggested that image quality should be defined in terms of a judgement of its suitability for its intended purpose - that quality is "task dependent" (Kundel 1979; 1990, conference paper, Nato ASI). For example, the image shown in Figure 8.1 is almost entirely lacking in contrast and detail, and yet brilliantly conveys the purpose of its creator. An interpretation of "unacceptable" in this context is that the displayed images should result in diagnostic accuracy at least no worse than that of the system it would replace - conventional film.

Subjective assessment of image quality is notoriously misleading. It is not sufficient to look at an image on a display system, observe a bony trabeculum or a subtle soft tissue feature, and conclude that image quality is therefore globally adequate. Nor is it enough to study radiologists' preferences, despite the fact that such studies still occasionally appear in the literature. The need for objective assessment is widely accepted.

Figure 8.1: Londres, le Parlement; trouée de soleil dans le brouillard. Claude Monet, 1904

In the hierarchy of tests of imaging system performance, physical tests can be used to study some of the most basic parameters. Such tests are unfortunately not necessarily predictive of performance with real clinical tasks. Phantom and simulation studies are also limited by the difficulty of relating test lesions to the clinical task, and model ideal observers. Clinical test series also have limitations: these are discussed elsewhere, but include the difficulty of reaching conclusions that extrapolate to pathological material not included in such studies, and the fact that participating observers have a heightened suspicion of the lesion under study. Prospective clinical studies would provide the ultimate means of evaluation, but are extremely difficult to perform and have not yet been attempted in more than a limited context. For the present, observer performance studies based on clinical test series provide the most realistic approach to objective evaluation.

An important issue in choosing clinical images for study is that it must be possible to create doubt in the mind of the observer, though an essential prerequisite is that the observer should be familiar with the condition under study. Measuring the certainty with which each observation is made enables more sensitive comparisons, and keeps the number of images that must be studied down to manageable proportions.

That is why conditions such as sub-periosteal resorption are frequently chosen for evaluating system performance,

though it is interesting to note that even with the most subtle lesions, radiologists incline to attempts at binary decisions in the majority of cases.

Both System 1 and System 2 were supplied as current, commercially available display systems by a major manufacturer, and were purchased as such. It had been considered that results from such systems were of more immediate relevance to the wider introduction of new technology than results with non-supported research prototypes.

8.2. SYSTEM 1

A significant loss of quality between film and all of the digitized displayed images was observed using this equipment.

It is worth noting that digitization on System 1 was to a higher resolution than was feasible on System 2. The weakest link, as far as image quality was concerned, was almost certainly the display monitor itself; the interlacing in particular, flicker, and other physical factors combined to degrade the resulting images.

A significant degradation was also associated with the compression algorithm that was incorporated.

Although only two clinical series were studied on this system, a confident conclusion was reached that the images were unacceptable for clinical diagnosis, a view that was accepted by the supplier. In combination with the other operational difficulties encountered in setting up the system, this resulted in recognition of the need to replace it.

8.3. SYSTEM 1 vs SYSTEM 2

The digitization parameters for the images displayed on the two systems were slightly different, on account of the different scanning spot size, and the fact that the scanning area for the digitizers on System 2 does not adjust to take account of smaller film sizes.

The matrix size resulting from digitization at full resolution of the hand images on System 2 therefore best approximates to "1K". It can be seen that the performance of these images is appreciably better than that of the "1K" images from System 1, and that these images are also better than the "2K" images from System 1.

These differences may reasonably be attributed to the improved performance of the monitors.

However the difference in performance of both systems, relative to original film radiographs, remains substantial.

8.4. SYSTEM 2

Improved specification of this System, improved technical support and generally improved reliability (though there were still problems with prolonged down-time) enabled much more thorough investigation than had been possible with System 1.

In many ways the most satisfactory group of clinical images investigated was the Pneumocystis carinii pneumonia series, largely because of the fact that non-radiological confirmation of the diagnosis had been obtained in every (abnormal) case. This is certainly one of the largest published pathologically-validated clinical series of its kind.

The results from this series seemed unexpectedly promising in relation to diagnosis from film, but require cautious interpretation. During the observation experiments, many of the radiologists appeared to be making judgements that were based not simply on perceptibility of the subtle interstitial shadowing that is typical of the condition, but that also took account of a general impression of increased "whiteness" of the lung fields, an impression that sometimes appeared to have been made before viewing any part of the image at full resolution with the "zoom" facility. In several instances, observers were disinclined to view the image with zoom, having already reached a decision on the basis of the lower resolution but full-

sized image. The issue of zoom has been discussed earlier (page 54) and requires more detailed investigation in its own right.

Irreversible data compression, which might be expected to reduce perceptibility of fine structural detail still further, resulted in a paradoxical slight improvement in performance over the non-compressed images, though not a statistically significant one. This might lend support to the view that general "whiteness" of the image, rather than perception of detail, had been a significant pointer to the diagnosis.

In this rather artificial situation, where observers knew that the diagnosis had to be either PCP or normality, the imaging system must nonetheless be permitted to take credit for a reasonable performance, particularly since no special image processing or enhancement techniques had been applied; how well this would extrapolate to clinical practice may be another matter.

As far as the mammograms were concerned, performance in regard to detection of microcalcification was extremely disappointing. It is interesting to speculate whether a higher resolution digitizer would have improved performance, but the most significant limitation to spatial resolution in System 2 is in fact the display system. The improved contrast resolution of the system proved unable to afford sufficient compensation.

Some might argue that these pathological entities with their corresponding subtle radiological features are of academic interest rather than of prime importance in front-line patient management. It would be more difficult to put forward such an argument in the case of the skull fracture series. Skull radiography is a commonly performed procedure, and the number of patients whose skull fractures would not be diagnosed correctly is indeed worrying, particularly in view of the fact that many gross example were included in the series. The findings in this series give rise to concern also about the implications for diagnosis of other subtle fractures.

8.5. DATA COMPRESSION

Data compression is an attractive and desirable solution to the problem of coping with the huge amounts of digital data that a hospital-wide PACS installation would be capable of generating.

When lossless data compression algorithms are applied, encoding and decoding yields data that is, by definition, identical to the original. Such algorithms now enable compression ratios in the range 2:1 to 3:1, and have no consequences in relation to image quality. Such an algorithm is built into System 2.

"Lossey" algorithms do not enable retrieval of identical data, however, and do require careful evaluation before

clinical use can be contemplated. During the period of this study, the effects upon clinical diagnosis of proprietary data compression algorithms supplied by manufacturers of two different commercial systems were examined. Both systems were "closed", in that it was not technically possible for us to apply our own choice of algorithm.

On System 1, the blocked discrete cosine transform algorithm, when applied to the "1K" images, resulted in characteristic block artifacts. The artifact remained visible with the "2K" compressed images. The ROC study showed a significant difference between the curve for "2K" images, and the curve for "2K" compressed images using this algorithm, and also a significant loss of quality in comparison with film.

Although the block artifact remained visible, it has yet to be demonstrated that such artifacts in themselves impair performance significantly.

System 2 incorporated DPCM with Lempel Ziv encoding. At maximum settings (level 1), approximate compression ratios were 8:1. Following lengthy negotiation with the manufacturer, we were permitted to use this.

With the PCP cases, the slight apparent improvement in performance of the compressed images has already been commented upon. With the digitized mammograms, the

reduction in performance following data compression was small, and not statistically significant.

As will be discussed later, there has been considerable progress with new data archiving systems, enabling higher storage capacity on low-cost media (such as optical tape). This may reduce some of the pressure to introduce need for data compression as far as storage is concerned; in respect of data communication, however, the need for data compression remains.

It is clear that a certain degree of concern exists about the possible medico-legal implications of implementing non-reversible compression algorithms. Such anxiety is misplaced, since the effect on image quality of some of the better algorithms that have been developed is probably much smaller than those of a host of other physical parameters that relate to the VDU displays themselves; this issue is considered further on page 249.

Our results with System 2 seem reasonable, and suggest that clinically useful data compression ratios can be achieved without major loss of image quality. Further work in this area will be essential.

8.6. METHODOLOGY

ROC studies are normally performed on binary decisions, normal/abnormal, or "signal" present/absent. However,

radiological reporting more typically involves making a statement about an image, or (in many cases) making a differential diagnosis, identifying several features and commenting about their position.

Ideally, observer performance studies should therefore include the ability to localize features, as well as modelling situations in which there are multiple, perhaps many, possible decision outcomes. The trouble is that such data are not easy to analyze, though a number of centres are working to extend ROC methodology to situations that are more representative of the clinical diagnostic process.

The methodological experiments considered here accepted the limitations imposed by the conventional ROC approach. Most of the performance studies that have been published previously have used pooled ROC data, though on theoretical grounds it seems clear that paired data methods are strictly more correct. This study was able to demonstrate that the use of paired data would not have altered the conclusions, though it might have increased the statistical significance of the results. A conclusion was also reached that it was unwise to encourage observers to increase their efforts to use the rating scales in a more uniform manner than they had already been doing.

8.7. PHYSICAL FACTORS

It is clear that the results from the clinical experiments show a considerable reduction in diagnostic performance for the displayed images when compared to the original films. Assessment of some of the physical factors that influence image quality suggest that these may have been an important element, and illustrate the need for careful testing.

Some of the physical factors that are potential sources of displayed image degradation are given in Table 9. On the left are listed the more fundamental parameters, and on the right are listed the more obvious types of visual effects that they cause. Although one ideally would wish to quantify the more fundamental parameters, one is in practice constrained to measure the more accessible parameters, indicated here as 'effects'.

TABLE 9: Physical parameters influencing displayed image quality

FUNDAMENTAL PARAMETERS:	EFFECTS:
Stationarity with position:	Vignetting Spatial distortion Changes of colour Grey level uniformity
Stationarity with time:	Flicker Jitter Reproducibility of grey scale
Signal to noise ratio:	No. of grey levels Electronic noise Video line visibility
Transfer function:	Resolution Veiling glare Sharpness of edges

'Stationarity' is that characteristic of the system that relates to the performance of the display at different positions across its surface. Many of the parameters listed here interact with one another: for example, signal to noise ratio is significantly affected by flicker.

Optimizing image capture and display - the input and output of images to and from a PACS network - is crucial, and without adequate quality PACS networks can serve no clinically useful function. This is not to say that problems associated with network management, communications between PACS components and the RIS or HIS, and other related technical issues are without difficulties themselves; as Gur (1989) has observed, optimization of subsystems does not equate with optimization of the entire process.

Data capture from CT scanners, MRI devices, and other digital systems is essentially an electronic and software problem - albeit one that is often complicated by manufacturer's reluctance to divulge what they consider to be proprietary information. On the other hand, defining the requirements for capture and display of radiographic images, and objective testing of the performance of such systems to establish their performance and to ensure that they meet those requirements, is essential in determining whether they can be used clinically. Much of this published material has so far been anecdotal.

Clinical and physical test data sets are required. Clinical data sets should represent the most difficult types of cases likely to be encountered, as well as cases more typical of the routine clinical workload. The physical tests performed should be capable of separating the performance of the components of the system, such as between the digitizer and the display. The required physical parameters are closely related to the type of clinical material under study: there is accordingly an interaction between the physical tests and test data sets, and the potential clinical uses of a system.

The behaviour of the human visual system makes certain types of error more significant than others. The eye is very insensitive to variations in grey level across the display, for example. Objective assessment of flicker (temporal intensity fluctuations) and jitter (temporal displacements) was important, however, since these appeared to be very disturbing with respect to the interpretation of difficult images, in particular on the display from System 1. It was felt that flicker had been a significant factor in the difference in performance observed between System 1 and System 2.

The precise effects of flicker cannot be easily determined. The visual system is not constant in its temporal response, and certain frequencies can be inhibited. In the retina, the rods do not respond to frequencies greater than about 20Hz, while the cones have a higher frequency cut-off. Thus

flicker affects particularly the peripheral vision. This is especially troublesome with arrays of monitors in the conventional layout of PACS consoles.

The overall results, with respect to the clinical data sets as well as the physical performance tests, indicate that great care needs to be taken when designing and setting up such systems if they are ever to be used for primary diagnosis. Further trials, using appropriate data sets, are a necessary part of the process of moving towards the implementation of clinically viable PACS networks.

To this end, there is now an urgent need for well-validated, standardized tests and test data sets, both clinical and physical. This will require a co-ordinated effort between different research groups, and should also involve manufacturers, but it will not be easy to accomplish. One problem, in the case of conventional radiographs, is that original films are needed - copies are not of sufficient quality; originals are easily damaged by handling and some digitizers cause scratches; the effort of generating test data sets is so great that most institutions are reluctant to release them.

Test data sets in the form of digital data, that can readily be transferred between systems, and can eliminate the influence of acquisition systems, must also be included.

9. PACS: THE WIDER ISSUES

9.1. INTRODUCTION[1]

It is now appropriate to consider some of the wider issues that relate to the context of digital display systems, and that will inevitably influence the timing and manner of their introduction into clinical practice.

9.2. IMAGE ACQUISITION

Conventional projection radiography, as stated earlier, represents roughly two thirds of the workload in most modern radiology departments: handling the resulting images represents the single most challenging technical issue in digital radiology.

In the quest for a solution, the current front-runner is computed radiography. There are believed to be more than 30 CR installations in the USA, more than 350 in Japan, and about 40 in Europe (October 1990). In some circumstances, CR is an end in itself. A hospital in the USA recently installed CR solely because about ten per cent of its ICU films were not being reported. No radiology report meant no revenue. Now the CR system is used to print out two copies of every image, which generates a worthwhile profit for the radiology department.

---

[1] A condensation of parts of this discussion was published as an editorial in Clinical Radiology, July 1990 (see Appendix 4).

Newer, cheaper, and more compact systems have been under development and are soon expected to become available so-called "desk-top CR"). This is likely to result in a substantial increase in use.

It is now emerging that the dose reductions so eagerly anticipated may not be substantial, other than in clearly defined circumstances. Witte (oral conference paper, Nato ASI 1990) has stated that in practice, there is no significant dose reduction associated with the use of CR in neonatal intensive care. This is not really surprising, since the X-ray absorbtion properties of CR plates are similar to those of intensifying screens; furthermore, the increased noise in a low-dose image results in an appearance that may resemble features of hyaline membrane disease, a common complication in neonates in the intensive care unit.

It is important to recognize that CR is not in fact an example of direct digital acquisition - the image is acquired as a shadow, an analogue 'picture' that is digitized by the scanning laser. This same consideration also applies to items of equipment such as the Konica Direct Digitizer (KDD) - a closed, dedicated chest radiography unit that incorporates a photostimulable phosphor.

Direct digital acquisition, as far as projection radiography is concerned, is not yet a realistic option,

though it is currently under active investigation. One promising approach involves the use of charge coupled devices (CCDs). CCDs are currently used, for example, in video cameras; they behave as arrays of tiny electronic detectors, each able to capture the information of a single "pixel"; the ideal radiographic CCD would need to be as large as a sheet of X-ray film - not achievable at present (the largest CCDs are currently 5.5cm x 5.5cm). Experiments with arrays of CCDs have met with only limited success; the junctional zone between adjacent chips yields unsatisfactory images. Large chips, or arrays of multiple chips, would lead to a further problem: a huge amount of data from every image (CCD pixel size is typically 25μm). Further experiments have used fibre optic minifiers to transfer full-sized images to CCDs (from phosphor screens): one application where this has been successful is in localization of impalpable breast lesions, where high resolution images have been generated and displayed in real time. Further possibilities include coupling scintillating fibre optic elements to CCDs or CCD arrays.

Film digitization has only a limited function within digital radiology. Sheets of film are scanned in a manner that is analogous to the laser scanning processes used in CR systems. The resulting digital data corresponds to a map of the optical density of the film rather than a map of stimulated luminescence. Film digitization is used principally in the research arena, as in our experiments, predominantly as a gateway for images that will be used in

evaluation studies. The effort of digitizing an existing film archive would only be justifiable in the context of a fully-functioning, more extensive digital network.

Film digitization involves conversion of the optical density distribution recorded on the film into a digital image array. Digitizers accomplish this by measuring the transmission of light through the film at discrete intervals. The intensity of light passing through is detected by a light-sensitive detector, and an analogue-to-digital convertor (ADC) quantizes the signal into digital image data. The film image is transformed into a two dimensional array of pixels.

Images derived from projection radiography carry a huge amount of information (See Table 2, page 43). Spatial resolution and the amount of data that result from digitization depend upon the digital matrix size; with scanning laser digitization, this is limited by the size of the scanning laser spot. The amount of data is also related to the number of grey levels represented for each point on the image matrix - the bit depth. The dynamic range of radiographic images is very large: more bits per pixel means better contrast resolution, but also more digital data that will need to be processed, displayed and stored.

TABLE 10: Amounts of image data associated with different radiological modalities. After Cox et al, 1986. Figures allow 1024 bytes allocated for header information.

MODALITY	BIT DEPTH	MATRIX SIZE	Mbytes data
Radionuclide imaging	8	128 x 128	0.017408
		64 x 64 (dynamic)	0.005120
Ultrasound	6	512 x 512	0.263168
CT Scanning	16 (acquisition)	320 (diameter)	0.161869
		512 (diameter)	0.412799
DSA	10	512 x 512	0.525312
		1024x1024	2.098176
MRI	10	512 x 512	0.526336

The sheer mass of data that X-ray departments are capable of producing represents a major problem. The London Stock Exchange generates 350 Mbytes settlement record data from each day's trading (personal communication, system supervisor, 1989). In contrast, a fully digital Radiology Department in a typical 500-bed hospital could easily produce 4GBytes image data in a single day - 12 times more (Glass & Slark, 1990).

In the modern X-ray department, many image types begin life as digital data - notably CT, MRI, and radionuclide scanning. Conversion of the images and signals produced by digital fluorography, DSA and ultrasound scanning, from analogue to digital form, is a routine electronic process that poses no particular problem. All of these image types have matrix parameters that are considerably smaller than those necessary for projection radiography (Table 10). Such images are in fact relatively easy to acquire, handle, display, archive and communicate in digital form.

9.3. DISPLAY

Kundel has wryly observed that the history of modern radiology is in fact a story of declining image quality (conference paper, Farwest Image Perception Conference, 1989): the best radiographs were produced on non-screen fine grain film; but accepted practice is to use rare earth intensifying screens and faster films with low exposures to

limit radiation dosage; and to use a broader focal spot in certain circumstances to preserve X-ray tube life or to reduce equipment costs. Radiologists take pride in their carefully nurtured diagnostic skills - observation, interpretation, and deduction; their ability to home in on the unexpected lesion or the subtle incidental finding sets them apart from their non-radiological colleagues. Radiological diagnosis depends upon image display. Is there scope for further compromise on image quality?

Images on film are usually referred to as "hard copy". Images on TV display systems are, to use the same jargon, "soft copy". Digital radiology in its fullest concept only makes sense if TV images can substantially replace the use of hard copy throughout the hospital, with "soft copy" reporting and viewing by radiologists and referring clinicians.

One index of the quality of images that can be displayed on TV systems is the number of raster lines on the TV monitor. Existing monitors used to display CT, MR, and DSA images are perfectly adequate, and usually display up to 625-lines. It is widely accepted that higher specification is necessary for satisfactory display of radiographic images. Higher specification, however, does not come cheaply.

Designing a suitable high resolution monitor is a costly and complex task. In order to achieve refresh rates that avoid flicker in the perceived image, some monitors use an

interlaced raster pattern - half of the lines are refreshed at each pass; the interlacing results in a movement effect that can be disturbing to the eye, and that many experts consider unacceptable. Other monitors seek to solve the problem of flicker by using a more persistent phosphor - sometimes causing a visible afterglow between images. There is a ten to twenty-fold difference in image brightness between film on a fluorescent light box and radiographic images displayed on a monitor. This results in reduced grey scale resolution for the displayed images and a critical need for dark viewing conditions; simply increasing the monitor brightness causes blooming of the phosphor, an increase in veiling glare and a degraded image - attempts to compensate for this have included the use of fibre optics in the monitor face-plate. The cathode ray tube is not an ideal instrument for the display of stationary, high resolution images, and research efforts are exploring alternatives: among the most promising appears to be the active-matrix liquid crystal system (Kazan 1989).

Most of the monitors currently proposed for PACS have 1024 or 1280 lines - none of the major system vendors currently offers a higher specification. 2048 line monitors have progressed from a research stage to a point where a small number of specialist companies are now able to supply them; many of the manufacture and design problems have been solved, but one current sticking point is their satisfactory integration into networks capable of handling

the huge amounts of data associated with each image; their high cost derives not just from the monitors themselves, but from the computing power often needed to support them.

A major PACS meeting (Medical Imaging IV, SPIE, February 1990) held a special seminar on high resolution displays, at which manufacturers and experts from many countries were present. A broad consensus was reached that 1024 and 1280 displays are inappropriate for the task of primary radiographic diagnosis in the majority of clinical settings, and that interlaced displays are also unsuitable. Even existing 2048-line monitors were considered not be adequate in their present form, despite the mounting pressure for clinical implementation.

Monitors of lesser specification than 1024 or 1280 lines have also been proposed, principally to permit review of images outside the main X-ray department ("secondary diagnosis"). At the same meeting, Kundel called this approach into question: can we really supply such images to our clinical colleagues on the understanding that they should not be used as a basis for important clinical decisions?

Objective assessment of image quality on high resolution displays is an issue of the utmost importance. Many of the physical measurements and methods that can be applied to an image recorded on a sheet of film have no direct counterpart. And there are many psychophysical differences

between the task of interpreting an image on film and a displayed image (Kundel 1985, Kosslyn 1989). Receiver Operating Characteristic (ROC) methods have been developed and refined to address some of the resulting problems (Metz, 1978, 1986, 1988, Swets & Pickett, 1986, Chakraborty 1989); they focus upon the diagnostic performance of the displayed image rather than any single physical parameter. ROC studies provide an objective means of comparing performance of different display systems with each other and against film. They have their drawbacks, however: they are difficult and time-consuming to run, and typical ROC tasks - such as determining the presence or absence of a pre-defined lesion from a series of similar images of proven pedigree - do not adequately simulate the real-life radiological task of making a difficult diagnosis from scratch. Subjective assessment of display quality - "eight out of ten radiologists found these images acceptable" - is a similar approach to that used in TV petfood commercials, and is not considered appropriate.

Abandoning hard copy reporting represents a fundamental change of practice that can only be justified after the most searching exploration of the consequences for radiological diagnosis. Such work is still in its infancy, partly because the technology has only recently reached a state of sufficient maturity to attract the active participation of diagnostic radiologists; objective assessment is an irksome process that cannot always be in

harmony with commercial pressures and consumer aspirations, but we cannot do without it.

Worldwide, almost every institution that has so far installed CR equipment still chooses to print out each image as hard copy. The CR hardware includes a laser printer for sheet film, and these hard copy digital images are at present used for primary diagnosis; this is not quite as wasteful as it seems: square foot for square foot, suppliers have so far held the cost of laser printer film to roughly that of X-ray film; the laser printer images are minified, however, so that the film costs of a CR hard copy service are indeed less. The fullest implementation of digital radiology would allow printing of digital images onto film only in special circumstances, such as for patients being referred to "non-digital" institutions; however, the "filmless hospital" may ultimately prove to be as inappropriate as a completely "paperless" office. (Johansson et al (1990) have recently reported promising results using continuous ink jet printers to produce low cost paper printouts from CR images.)

9.4. IMAGE PROCESSING

Once in digital form, data processing techniques can be applied to enhance, manipulate and interrogate the resulting images. Such processing should be distinguished from simple manipulation of window level and width. Examples include techniques such as subtraction, edge

enhancement, noise reduction, contrast equalization, and dual energy filtration, but experience suggests that enhancement of one feature is often at the expense of another (Oestmann et al, 1989). Existing TV display systems inevitably introduce image degradation: it remains to be demonstrated whether image processing methods and window manipulation can adequately compensate for the loss of quality, and can indeed contribute to improved diagnosis.

9.5. DATA STORAGE

As a data storage medium, film is costly, bulky, and requires manual filing and retrieval - a tedious process that is notoriously vulnerable to error. What is more, a sheet of film can only be in a single location at any one time. In a single exposure, the amount of information that can be captured is restricted by the latitude of the film. Such limitations have provided powerful impetus to the development of digital systems.

Magnetic hard disks allow rapid access times, but have a limited capacity; they are costly, but data can be continually over-written; their most useful rôle will ultimately be the provision of rapid-access buffer storage rather than a permanent archive. In December 1989, IBM announced a new method of data storage, increasing the density of magnetic hard disk storage by a factor of 15-30 with no loss of speed (product announcement, Dr K Keeshan,

IBM, San Jose, USA). This may have important future implications for PACS.

Optical disks store digital data at higher density. They are relatively cheaper, and a little slower. Storage is permanent, and the data cannot readily be over-written. Fourteen-inch conventional disks can store 6 Gbytes per disk, though this is still not enough to enable single disks to provide adequate on-line storage: optical jukeboxes have therefore been devised to handle 90 or more disks at a time. Widespread application of compact disk technology in the domestic audio field has so far provided the greatest stimulus to development and refinement of optical disk technology, and the manufacturing costs of compact disk drives are now small. It remains to be seen whether there will be any benefits in terms of improved design and reduced cost of optical jukeboxes, which are at present cumbersome, mechanical and unwieldy. Current models can cost as much as £1/2 Million to install; "mirroring" data, for extra security, may necessitate a second, backup archive - and some would deem this an essential requirement.

There has been a recent new development that appears to hold promise - the only British contribution to the technological scene. It is called optical tape, and has been developed in the UK by ICI (Pountain 1989). It uses optical storage technology with higher capacity at a fraction of the cost. A single twelve-inch reel of optical

tape is capable of storing 1 Terabyte of data - a million megabytes. Access times are relatively slow - 58 seconds from one end of a Tbyte tape to the other (and proportionately less with shorter tapes). The film cost of a chest X-ray is about £1.50 per sheet; storing the same image on optical tape costs about 1.5 pence. The first optical tape drives (CREO Systems) are currently under evaluation in Canada, and are planned for supply in Europe during 1991 (personal communication, D Bennett, ICI Imagedata, UK). Drive cost is likely to be around £120,000, but a viable archive would need more than one. The prospect of being able to set up a number of duplicated image archives at various points on a PACS network is alluring, but it remains to be seen whether this medium will truly meet the high expectations that are currently held for it.

Other storage media are also worthy of mention. Optical tape technology may also be applied to the development of optical floppy disks (Bernoulli Systems), capable of storing 1 Gbyte or more at low cost. Magneto-optical disks are more expensive, but are re-writable; a 5.25 inch disk can store 660 Mbytes. A credit card sized piece of optical tape would be capable of storing between 150 and 300 Mbytes; at a cost of pennies, it could reasonably be issued to patients as a portable, back-up copy of their own images. A prototype system based on 3.5 inch WORM compact optical discs (capacity 180 Mbytes, and more expensive) has already been successfully introduced on a small scale at

UCLA to do just this (Cho et al, 1990), and 3.5 inch magneto-optical disks are also being investigated for this purpose. Such developments would also make it possible to download copies of image data for clinics or selected departments, relieving demand on the main network, or perhaps reducing the need to extend the network far from the radiology department.

Storage capacity can be increased by means of data compression techniques. It is possible to encode data with no loss of information at compression ratios of up to about 3:1. Further compression may result in slight data loss, but it seems likely that clinically useful compression ratios of up to 20:1 may be feasible. These so-called "lossey" algorithms require careful evaluation for diagnostic image quality using ROC techniques. A number of manufacturers, notably in the USA, have expressed deep concern about the medico-legal implications of wilfully applying techniques known to impair image quality, and this is discussed further below.

9.6. NETWORK FUNCTIONS

Connecting up individual system components is the easiest task in the world, but only in theory: all it takes is a few pen strokes on a diagrammatic plan. In reality, computerized networks are complex, fragile systems that demand the utmost care in setting up and supporting. The network problems have absorbed much of the research effort

that has so far been invested in PACS, both by manufacturers and by individual institutions (Templeton et al, 1988).

Linking and interfacing different types of equipment, and finding ways for existing and future equipment to communicate in the same language is a complex task. An interface is a device that connects two independent systems and enables them to interact. The physical connection established between devices by an interface must conform to certain connectivity standards, and the nature of the interaction between devices is specified by the interface protocol. There are three types of interfaces encountered in image management systems.

1. Imaging equipment interfaces, linking imaging systems to network acquisition and formatting nodes, providing for the access of the digital image data generated by the equipment.

2. Network interface units, linking the network nodes to the network transmission medium, enabling the transfer of digital imaging data to and from a local area network.

3. Gateway interface units, linking the local image management network to more generalized information management systems.

A standard is a set of definitions, rules and requirements concerning classification of components, specification of materials, performance or operations; and or measurements of quantity and quality. Conformity to standards should in theory assure that a particular piece of equipment or software will be compatible with a variety of systems. Adoption of a standard allows products from multiple manufacturers to interface with each other, allowing the purchaser flexibility in equipment selection and use. The purpose of standards should be to harmonize, not restrict. The disadvantage is that, once established, they tend to inhibit further research and development. Furthermore, by the time a standard has been developed, evaluated and implemented, more efficient techniques are often available or under investigation.

The ACR-NEMA digital interface was developed by a joint committee of the American College of Radiology and the National Electrical Manufacturers Association. NEMA is essentially a trade organization, and a factor that has increased the complexity of the task was the constraint upon it that any agreed standard should avoid conferring a commercial advantage on any particular manufacturer.

SPI (Standard Product Interconnect) is an extension to ACR-NEMA developed by Siemens & Philips, in joint work since 1984. It has never been published, and is not being supported by other manufacturers.

The Medical Image Processing System (MIPS) committee is the Japanese standards organization set up in 1985 to address similar issues to ACR-NEMA. It has developed standards that can be regarded as a variant of ACR-NEMA using 16-bit characters (to accommodate kanji rather than ASCII). It is being amended in direction of ACR-NEMA.

Standards such as these are designed to overcome some of the connectivity problems, and are at an advanced stage of development (Horii et al 1990). They are costly and it is likely to be a long time before any manufacturer offers them as standard; the current generation of imaging devices was not designed with networking in mind, and networking can reduce or slow their performance; and they certainly do not yet provide a complete solution.

The primary objective of the ACR-NEMA set of standards is to enable point-to-point communication between devices. It does not lay down file format standards, for example. It has therefore been left to individual researchers to develop tools to ease the portability of image data across systems (e.g. PAPYRUS - Ratib et al 1990).

The hardware component includes a standardized 50-pin connector, used for the ACR-NEMA cable. Data transfer takes place at rates approaching 8 Mbytes per second. Commercial versions of the ACR-NEMA interface have begun to appear, and are offered by manufacturers of imaging devices as part of the imaging system. Laser printer manufacturers

also offer the ACR-NEMA interface for use with their devices.

Further problems are emerging. At a recent international meeting (EUROPACS '90), it became clear that there is a widely felt undercurrent of dissatisfaction with some of the shortcomings of ACR-NEMA, and that numerous institutions and companies have begun to make their own modifications, a situation that may defeat the purpose of having any standard at all. More recently, Chan et al (1990) have stated: "It is likely and hopeful that the 50-pin ACR-NEMA plug standard will be abandoned in favor of more modern, versatile, and cost effective network-style interfaces" ! The ACR-NEMA standard is currently under revision.

In a fully operational digital hospital, the network would have to be able to satisfy requests to distribute images belonging to any patient, to any location in the hospital: this would mean that thousands of images from hundreds of patients would need to be instantly accessible for viewing. The data traffic problems and the practical problems of data management, are enormous, though they are not necessarily the same in every country: in Sweden, for example, accepted procedure is that images do not leave the X-ray department; a Swedish network would presumably be easier to run than one at St Mary's Hospital, London, where it was at one time considered that a total system would require 140 separate image display workstations. A further

problem is speed: the amounts of data involved are potentially huge, and the systems so far available are disappointingly slow (Table 11). Data compression speeds data transmission, and is one partial solution; fibre optic cable links on a network can further increase speed - though they add further to system costs. Experience with a recent clinical installation in Italy designed to handle CT images, shows that reporting times have increased by more than 30 per cent, to a point where radiologists feel a powerful disincentive to use it (Ukovich et al, 1990). As noted in Table 8 (page 174), our own findings show typically a three to four-fold increase in reporting times for displayed images compared with film radiographs, with well-trained observers performing identical tasks. Ultimately, no radiologist will agree to work with a system that intrudes significantly upon his or her productivity, regardless how useful the images might be.

Another obstacle is the problem of linking text to the images through the radiology information system (RIS), and providing adequate communication with hospital information system (HIS) and patient databases. For example, the HIS will have an advance record of all patients attending a particular clinic; the RIS will have a record of those patients who have had radiological investigations; collating and transmitting the relevant images and reports - perhaps to the clinic's own database - should be an automatic and effortless process performed at "off-peak" times.

TABLE 11: Maximum data rates of network media. FDDI and B-ISDN are costly and not yet widely available.

Data transmission rates:	
ISDN	0.1 Mbits/sec
Ethernet	10 Mbits/sec
Token ring	16 Mbits/sec
FDDI	100 Mbits/sec
B-ISDN	140 Mbits/sec

Note: typical throughput is of the order of only ten per cent of these figures.

A fully functioning system offers the prospect of image communication and transmission not just across a hospital network, but between different centres - teleradiology. This is of particular benefit where distances are great and specialist skills are scarce. Teleradiology provides much of the impetus to PACS research in many areas of the world, including Australia, Scandinavia and parts of the USA, and substantial progress is being made towards implementation (Batnitzky et al 1990). At a number of centres in the USA, for example, neuroradiologists are already able to monitor patients undergoing CT scanning at clinics many miles away, and some centres even have their own satellite links. Transmission over the public telephone system is relatively slow and susceptible to degradation; while it is unsuitable for a large volume of image data, CT and MR images can be transmitted quite adequately by this route. Perhaps of more immediate practical importance in the UK, radiological reports are already being transmitted to referring general practitioners' own computers over the phone system.

Data security is an important preoccupation in business and industry. In networks such as airline reservation systems, it has taken years to refine. Data must be protected from unauthorized access and misuse as well as from inadvertent corruption, and in PACS too this will inevitably demand careful consideration.

9.7. QUALITY ASSURANCE

Quality assurance (QA) in some form will be crucial to successful introduction and acceptance of PACS, and will need to be applied to each point within a network.

Current methods of QA are the product of almost a century of evolution. Though tedious, QA is a necessity for every good X-ray department. Digital technology brings new and more complex problems of QA that have so far received little attention.

The Royal College of Radiologists defines QA as:

> **"A prospective programme of testing processes designed to ensure adherence to defined standards of quality."**

The World Health Organization definition is:

> **"An organized effort by the staff operating a facility to ensure that the diagnostic images produced by the facility are of sufficiently high quality so that they consistently provide adequate diagnostic information at the lowest possible cost and with the least possible exposure of the patient to radiation."**

For PACS, standards have not yet been agreed or defined: the necessary steps merge imperceptibly with acceptance tests, tests to ensure that performance is as specified, general housekeeping, and audit. QA may not be the best

term for this and assumes "quality" already exists. The FDA is considering terms like "performance maintenance". Anyone buying equipment at the cutting edge of new technology must accept that manufacturers may themselves be unfamiliar with necessary measures, and must be prepared to assume responsibility for establishing and maintaining appropriate programmes themselves.

Two main categories of equipment test need to be considered:

i) Overall tests, on the performance of the entire system.

ii) Component tests, to isolate single components, though in any network, it may be difficult to pinpoint the site of any loss of 'quality'.

There is a need both for clinical and physical tests, and each component has both hardware and software functions.

A PACS installation can logically be divided into the following components:

a) data acquisition (e.g. CR, film digitizer, frame grabber, digital link)
b) display workstation
c) archive(s)
d) communication channels
e) links to other systems (RIS, HIS)
f) hard copy devices

9.7.1. Data acquisition

Acquisition can be digital or analogue. Digital data capture from an MR system is purely digital. A film digitizer or a CR system is purely analogue. Different tests are required.

9.7.1.1. Digital devices

Digital data as captured should be identical to that on the original system. A simple test would be to return it to the original system for subtraction, or to compare actual numerical pixel values.

Other tests are needed for communication channels, and performance - especially in the presence of load. Patient identification and associated data must also be verified.

9.7.1.2. Analogue interfaces

Analogue interfaces include digitizers, CR, TV cameras, and frame-grabbers. Physical tests would include tests of resolution, signal to noise ratio, uniformity and stability.

To test such systems it is necessary to have a means of access to the data in the first place, which may require

more cooperation than many manufacturers are willing to provide.

Experience of testing acquisition devices in conjunction with the present investigation has been confined to laser film digitizers, which provide the gateway to the display system, and some of the tests undertaken have been described above.

It is important to recognize that the image plate reader in CR systems is also a form of digitizer, and that similar issues apply.

CR yields images of superficially similar appearance from a wide range of different exposures. In fact, images that are underexposed will have a reduced signal to noise ratio. The temptation to shoot off a series of images - on the ICU, for example - without modifying the radiographic factors between patients, can be overwhelming, and does happen (D. Perkins, Ochsner Clinic, personal communication). Preventive maintenance and general housekeeping are important. Frequent densitometry is needed. CR plates are re-usable, and quickly become grubby. They are easily damaged, and in practice may have a life-span as short as one year. The trouble-shooting and technical skills necessary to solve problems exceed what can be expected from unsupported junior radiographers running an out-of-hours service, and provision for appropriate in-house technical backup is essential.

9.7.2. Display workstations

Many of the procedures referred to in the sections on physical experiments and the discussion of physical factors, should be included in a continuing programme to ensure that performance remains as specified. Roehrig et al (1990) have proposed a daily programme of testing, based on simple observations of test patterns such as SMPTE, and the kind of simple equipment employed in the physical experiments that have formed part of this investigation; most of the observations do not require quantitative measurements, but merely the determination of the presence or absence of features under test.

Adjacent monitors should be carefully matched. Variation in the colour of the phosphor disturbs the eye even when within manufacturers' tolerances. This causes problems when a single monitor in a bank needs replacement. At St Mary's, it was necessary to replace original monitors with ones from a uniform batch; and to obtain specialist help to ensure that colours and brightness levels were identical.

9.7.3. Other components

These include the computer system, archive and communication channels; hardcopy devices; and links to other systems.

Performance testing of these components requires specialized skills, and requires careful provision: they are largely outside the scope of experience at St Mary's.

On the subject of archiving, however, the grubby disk shown in Figure 9.1 caused an intermittent archiving fault that baffled service engineers from four countries over a period of 4 months, and was discovered purely by chance.

9.7.4. Environmental conditions

Suitable operating conditions are essential, and need surveillance. System 2 suffered damage from temperature fluctuation, and dedicated air-conditioning should be provided.

Computer systems are noisy. System 2 generates 68 dB - enough to be a nuisance. Ideally, computer equipment should have its own room, away from workstations.

Ambient lighting is an important issue in PACS. Fluorescent lighting should be excluded from all places in which display systems are likely to be used. It accentuates flicker in the perceived image - especially in monitors running at 60 Hz. The brightness of all lighting should be easily adjustable, with appropriate provision to exclude daylight, and to avoid reflected glare from the monitor face-plate.

Figure 9.1: 12-inch optical disk, showing accumulated London grime. The disk has been removed from its cassette, and the lower half has been cleaned.

Surrounding equipment may also interfere. It took several weeks and many service visits, to discover that the unstable image on one of the monitors in System 2 was due to interference from an ordinary dot matrix computer printer on a shelf above it (the printer on the right in Figure 3.4 - page 94).

9.7.5. Total performance tests

Performance measurements range from ROC studies to parameters like time to display an image or sequence of images, time to rotate or move an image, time to zoom, window and threshold performance, and ergonomics.

ROC methods focus on the end result - diagnosis - rather than individual parameters. As implementation proceeds, emphasis will gradually shift towards prospective audit of diagnostic outcome.

In conclusion, QA is the responsibility of individual departments, not suppliers, and will remain so. There are 30-40 UK centres currently examining the prospects for introducing PACS. In the rush for implementation, the unglamorous task of performance testing and surveillance - and the staff and equipment to do it - should not be overlooked; in their plan for a total system implementation, Glass & Slark (1990) make no reference to it.

Current QA tests are easy to do, cheap to perform, are not time-consuming, and are easy to reproduce, which is where their sophistication lies.

9.8. WORK PRACTICES

What will be the effect of digital radiology upon the way we work? How will the relationship between radiologists and clinicians change? Should each image become accessible on the network as soon as it is taken, or only after it has been reported? Image transmission for case conferences between different institutions - perhaps even in different countries - could become commonplace. Clinicians dissatisfied with a radiological opinion could conceivably shop around by satellite...

Cho (1988) noted that a digital link to a coronary care unit resulted in greatly reduced personal contact between radiologists and clinicians. Balter (1989) has also cautioned about the impending "sterility of communication" that may follow widespread introduction of digital technology, if we are not careful; non-verbal communication plays an important part in the relationship between radiologist and their colleagues, conveying confidence, concern, and a feeling for how important or relevant particular radiological findings may actually be.

While it is intriguing to speculate on some of these issues, it may also be of limited value - much will have to

depend on eventual system configuration, performance and practical operating constraints.

9.9. MEDICO-LEGAL IMPLICATIONS OF DIGITAL RADIOLOGY[1]

9.9.1. Litigation in conventional practice

Medical litigation, worldwide, has increased dramatically. Records from the Forties and Fifties in the UK showed only small numbers of cases, to which radiology contributed a barely significant proportion. Today in the UK, radiology still only constitutes a small fraction; but in the USA radiology ranks third among medical specialties, and one in every five radiologists is sued for malpractice annually (Berlin 1984, 1986; Spring et al 1986). Patterns of malpractice lawsuits in the USA commonly anticipate trends in the UK and Europe: current trends are for a continuing, sustained increase, and changes in the pattern of cases.

In a retrospective review of 360 cases of radiological litigation from the records of the Medical Protection Society, London, spanning approximately six years, 78 per cent of the cases reviewed related to trauma in which an incorrect diagnosis had been made (Craig, Knox Lecture, Royal College of Radiologists, 1987). Initial radiographs had not been seen by a radiologist in 32 per cent of these cases, and no written report had been recorded. The most commonly missed fractures were of the scaphoid, radial head

---

[1] I am grateful to Dr J O M C Craig for his assistance with this section.

and femoral neck. The most commonly missed dislocations were of the shoulder (posterior dislocation) and the metatarsus (Lisfranc). Many of these injuries - though not all - were associated with the presence of subtle radiographic abnormalities that had been missed. The remaining 22 per cent of cases, unrelated to trauma, resulted mainly from errors in observation, interpretation and diagnosis. Examples include missed diagnoses on barium examinations, missed tumours on myelography and CT scans, and missed fetal abnormalities on ultrasound examinations.

The ratio between trauma and non-trauma cases appears to be changing in the USA, with a greater increase in the latter category. We may presumably expect to see similar changes in Europe. Litigation also arises from errors and accidents of a technical nature, which must inevitably increase with the expansion of vascular and interventional procedures, which will be largely unrelated to digital technology.

9.9.1.1. Current standards

The standard of practice in the UK for medical negligence was laid down in Law by Chief Justice Tindal, more than a century ago, who said:

> **"Every person who enters into a learned profession, undertakes to bring to the exercise of it a reasonable degree of care and skill. He does not undertake, if he is an attorney, that at all events you shall gain your case,**

> **nor does a surgeon undertake that he will perform a cure, nor does he undertake to use the highest degree of skill. There may be persons, who have had higher education and greater advantage than he has, but he undertakes to bring a fair, reasonable and competent degree of skill".**

In other words, to avoid a finding of negligence, practitioners are required to attain a reasonable standard of care, judged according to standards of practitioners of similar skills and expertise, practising conscientiously. More recently, Viscount Hanworth, in a speech to the House of Lords, said:

> **"It would appear that damages against our medical profession are rising and that we are in danger of following the American precedent of widening the meaning of negligence. In America, this now seems to be simply equated with a wrong diagnosis, or giving treatment which with hindsight was not optimal."**

As the use of digital technology increases, we are entering uncharted seas regarding litigation. Economic factors are important in relation to PACS: it is necessary to bear in mind that substantial costs may arise from errors attributed (deservedly or otherwise) to inadequate systems.

9.9.2. Digital radiology: likely medicolegal issues

9.9.2.1. Soft copy reporting

Among the most important issues are those concerning image quality and the acceptability of displayed images for primary diagnosis.

There are certainly significant differences in diagnostic quality between the displayed digitized images and the original film radiographs in this study. There are grounds for concluding that systems of the type studied may be unsuitable for the diagnosis of subtle lesions.

The most acceptable array - "1K", "2K", "4K", or even more - remains undefined. Can the acceptable acquisition and display matrix sizes vary according to the pathological diagnosis - even though the true diagnosis may not be known at the time the examination is being conducted? Should matrix size vary according to the body part under examination - just as we currently reserve high resolution techniques for certain types of extremity radiography, for example? It would not be sufficient to argue that a particular matrix size or image quality is acceptable in the majority of situations: when litigation occurs, it is only the case in hand that matters.

Diagnosis from remote viewing stations within a hospital is also a troublesome issue - especially if these are of

inferior specification to those in the radiology department. If the image is not accompanied by a radiological report - at night, for example - the referring clinician may be left to make a diagnosis from an inferior image. If an error is made, it is unclear who would bear liability.

Batnitzky et al (1990) have also raised the question of liability in relation to teleradiology: who bears liability in relation to failure of diagnosis from a transmitted image?

9.9.2.2. Provision of reports

There must be appropriate safeguards to ensure that each image on a network is accompanied by a report, that the correct report is provided with each image, that each image on the system has been reported, and that each report has been reviewed by the referring physician. Given present levels of radiological manpower, however, it is uncertain that the first of these objectives will be consistently achievable.

In the UK, information stored on computer systems is subject to many legal requirements, including disclosure. Confidentiality and security must be carefully maintained, and if it is neglected, a legal liability exists. It is not clear at present whether this rests with manufacturers or with operators, or both.

9.9.2.3. Data compression

It remains unclear whether it is legally acceptable to compress images for more efficient storage and transmission on the network, and what kind of compression ratios would be acceptable. Any wilful loss of information from data compression may be culpable. Manufacturers have focused much of their concern regarding the medico-legal consequences of digital radiology upon this single issue, and the anxiety of the manufacturers of System 2 in this regard has already been alluded to. However, such considerations might equally apply to many other aspects of this evolving technology that result in images of lesser resolution than is currently achievable with film, and the effects upon image quality of low levels of data compression are likely to be considerably less significant than factors like the display itself. The experiments in this study have shown that incompletely reversible compression ratios of 7:1 have had little effect upon displayed image quality, but it is likely that manufacturers will none-the-less remain shy of all but lossless algorithms.

With the growth in digital techniques such as digital subtraction angiography and digital fluorography, vast numbers of images can be generated. A most basic compression exercise is to decide which images to store and which to discard. The legal consequences attached to this first "compression" process remain also largely unexplored.

Plans to introduce data compression for archived images are currently an integral part of the Madigan project (MDIS 1990): a compression ratio of 10:1 will be applied after reporting, and the administrators of the project have termed the process "digital microfilm".

9.9.2.4. Failure to maintain, upgrade or install state-of-the-art equipment

There may be legal implications if reports or images are lost or are unavailable through technical error or equipment failure, and a patient suffers. The same considerations apply to accidents and delays through "down time".

It may furthermore be legally indefensible to fail to upgrade installed digital technology, as developments continue to take place. A precedent perhaps already exists in respect of conventional technology - a London teaching hospital was once involved in litigation because, in the absence of a CT scanner, a carotid arteriogram had been performed; the patient developed a hemiplegia, and the hospital had to accept liability for failing to offer a modern CT service.

Equipment modification may also pose a problem, especially in relation to changes to software by users - which may sometimes be essential.

Manufacturers are subject to the laws of product liability, and will seek advice from their own lawyers in deciding exactly what equipment they feel able to bring to the market place. (The situation could change: in the USA, if the FDA were to decide that PACS components fell within its jurisdiction - by declaring PACS to be a medical "device" - the onus would shift firmly towards manufacturers to prove safety and efficacy, as with a new pharmaceutical product.) However, standards of clinical acceptability can only be set by the medical profession itself, which must decide what is diagnostically safe. This must be on the basis of scientific evaluation and ongoing clinical experience. What might be acceptable in 1991 may not be acceptable even a short time later and the evolutionary nature of this technology may itself create legal problems when cases come to trial years later.

Most of the medico-legal issues arising from digital radiology remain unanswered questions; as these standards continue to evolve, the clinical acceptability of this technology will change and will require ongoing examination regarding its legal credibility.

9.10. IMPLEMENTATION STRATEGY

It remains true that no institution in the world has yet achieved the target of a functioning, all-digital radiology department, and it is at present difficult to feel confident about the clinical performance of a total system

that would have to be constructed from currently available equipment. Even in Japan, where no fewer than 96 hospitals around the country are experimenting with PACS components beyond stand-alone CR installations (June 1990), and prototype networks have been established in 12 hospitals, only a single institution (Hokkaido University) is attempting a hospital-wide system (Okabe 1990).

Some may find this disappointing. After all, digital radiology has now been under development for almost a decade, and the original concept seemed simple enough; its visible progress has hardly matched the brisk advances that have taken place in other areas of medical imaging over roughly the same period. If there is an element of failure implicit in its slow progress, it is above all the initial failure to fully recognize the sheer technical enormity of the task that was being addressed.

Some may also find it difficult to be dispassionate about the need for implementation - there is an overwhelming temptation to believe that digital systems will somehow provide a blanket solution to all of the daily frustrations that go with providing a radiology service - particularly in the UK's overstretched and chronically under-funded hospitals - including constant trouble with lost or missing films, and with film filing, handling and administration.

Huang (1990) has observed that there are in practice three possible routes towards PACS implementation: a "home-made"

multi-vendor system requiring a large amount of in-house support; a customized system in which a vendor responds to a detailed specification tailored to meet exact needs; and a turnkey operation in which a vendor provides a complete system in an "off-the-shelf" package.

It **is** clear that there is a powerful and irreversible movement of modern medical imaging technology towards digital systems, but for the present it is by no means certain that implementation in a single step is essential or desirable, even on theoretical grounds. The technology is now hardly in an optimal state and is currently undergoing rapid evolution. This is a necessary, ongoing process, that can only be exploited to full advantage by a rolling programme of implementation that would allow networks to grow over time, with gradual redeployment of obsolescent equipment in an imaginative manner away from the most sensitive points on the network (Dawood 1989).

This viewpoint appears to be a controversial one, and the issue of whether or not total digital systems should be implemented in a single move has attracted a disproportionate share of the PACS debate in Britain, for two main reasons. Most radiology departments do not have a programme of capital equipment expenditure able to meet conventional replacement needs, let alone the cost of new digital systems; it is much easier to obtain funding for new equipment by incorporating the costs into the considerably larger construction and commissioning budgets

of a new hospital or wing. Secondly, economic justification is perceived by some to be essential, and the most impressive predicted savings only appear when calculations allow total elimination of costs that attend the use of film.

In Britain, hospitals currently planning for redevelopment therefore face perplexing choices, apparently between two incompatible extremes. The opportunity to "go digital" is viewed as one that may not come again, but is attended by a host of unknown quantities - technical viability and clinical effectiveness among them - that have to be reconciled with the need to provide a good clinical service from the first day of operation. Alternatively, failure to grasp this challenge may mean confinement to conventional, film-based systems with no further prospect of introducing digital technology until the day that the newly built department eventually falls down.

It is not the purpose here to seek to justify PACS in terms of cost; all kinds of computer models and statistical manipulations have been proposed to do so, but there are too many uncertainties to make any meaningful conclusion possible. The idea that a filmless system is necessarily a cheaper one seems fanciful - except perhaps in the very long term. It has long been realised that low-paid film filing clerks will have to be replaced by highly skilled computer staff and medical physicists (Gray et al, 1984). There will always be a need to produce certain images as

hard copy - no-one knows how many. Digital storage media - disks or tape - have a significant cost; and of course the digital equipment itself has a high capital cost, over which obsolescence looms almost from the day that the order is placed. Equally, it is unreasonable to judge the benefits purely in economic terms; some would argue that the prospect of a smoothly running and fully effective, integrated electronic X-ray department, with reduced radiation doses for patients and staff, is beyond price.

Nor is it intended to argue the case for proceeding with implementation: the merits of PACS have been reviewed quite adequately by other authors (e.g. Capp et al, 1985, Huang 1987, Craig 1985).

If PACS is to proceed in the UK, however, it will need to be properly funded. The National Health Service has so far been insulated from many of the true costs of modern imaging developments: CT and MR scanners that in other countries are the accepted responsibility of the prevailing health care system, have had to be obtained by appealing to public charity. PACS cannot be funded by public appeal: it is not one single item of hardware, but an investment in the fabric of hospital infrastructure, that will require a continuing commitment to support and maintain, as well as a more innovative approach to funding - which should permit implementation at a rate that reflects current progress and success with each stage.

9.11. PRACTICAL IMPLICATIONS

What can a hospital that aspires to embark on PACS implementation do now? Acquisition techniques in general radiography - such as CR - are likely to remain compatible with existing and future general radiographic equipment, so that no special provision needs to be made in this respect. New CT, MR, digital fluorography and DSA installations should be specified to incorporate (or be readily adaptable to) the latest version of standards such as ACR-NEMA and SPI. A number of institutions have already begun to establish small-scale networks to handle these images, with promising results. This is an appealing approach: image quality issues are largely irrelevant, and data management problems are on a different scale. The exercise serves as a worthwhile prototype of the larger task that lies ahead.

CR is now sufficiently mature to justify controlled implementation for further study and evaluation in areas where the benefit is likely to be greatest, such as for portable work and in the ICU; hard copy output will be necessary, but it would be worth installing a display system on a small scale for evaluation.

No less important is preparation of staff: continuing education and dialogue with others working with the same technology is essential. The objective must be to build up an infrastructure of supporting staff - especially radiographers and medical physicists - with the necessary

skills to implement larger systems successfully in due course. It is also important to correct operational inefficiencies: PACS will not cure poor motivation, sloppy administrative practices and bad technique - it may make them worse - and it will certainly bring problems of its own that have not yet been thought of.

New buildings should allow for ducting, air conditioning, water supply and drainage, power requirements, computer flooring, tungsten rather than fluorescent lighting, and provision of blinds. They should also have a layout able to cope with the co-existing needs of conventional and digital systems as implementation progresses. Choice of equipment from a constantly expanding pool is difficult - manufacturers are persuasive, and anxious to recover their substantial development costs; staff within a department should acquire the skills necessary to assess proposals objectively.

There are at present three major PACS projects currently aimed at achieving total system implementation. The project at the University of Hokkaido has in fact been under way for several years and has absorbed the efforts of several major manufacturers working together; implementation is in fact taking place in an incremental manner. The Madigan Army hospital, Washington State, is a new hospital now planned to open in mid-1991; progression to total digital radiology is intended to take place over a 6 month to one year period (MDIS 1990). The administrators

of the project consider that the US Army has unique requirements that justify such an experiment, including the need for teleradiology, battlefield support, support for disaster relief teams, and absence of a need to provide a long term archive for troops in combat; a substantial amount of funding has been set aside for the project, including provision for upgrading the system as developments occur. The project is currently out to tender. A further project is in progress in Vienna, at the 1400-bed Sozialmedizinisches Zentrum Ost (SMZO), where a new hospital is planned for completion in early 1992; a vendor has already been selected, and allowance has been made to upgrade the system over a 5-year period (Mosser et al 1990).

In the UK, a proposed total system implementation is planned at the Hammersmith Hospital; it is scheduled for late 1991, and its specifications are based on those for the MDIS project (Glass & Slark 1990). There are proposals to follow this with an implementation at the new Westminster/St Stephen's hospital, due to be rebuilt by 1992, and several other sites in the UK have also declared their interest. Other UK projects include a CR installation in Leeds, and a multi-modality image network at Guy's Hospital.

Worldwide, the institutions that have so far made most progress towards implementation have however been established departments that have introduced new technology

step by step. They have not been purpose-built. They fall into Huang's first category (above), and have ongoing programmes of research and evaluation. They have relied upon no single manufacturer or supplier as exclusive sources of hardware or of know-how, and have devoted considerable effort to developing support skills in-house: at UCLA, for example, the digital unit now has a physics and scientific staff of almost 40 people. They have carefully cultivated links with academic and scientific departments at nearby universities. And they have generally had access to government funding and research grants.

10. CONCLUSIONS

CONCLUSIONS

The principal conclusions of this investigation may be summarized as follows:

1. Of the equipment studied, System 2 performed best. This configuration used 210μm pixel data presented on 1280-line display monitors. The results achieved, however, indicate that this system would be unacceptable for primary radiological diagnosis of subtle lesions. This conclusion has important implications for institutions that are currently considering introduction of systems based on similar parameters.

2. The minimum acceptable parameters for digitization and display of radiographs have yet to be defined by objective means; it is clear from the literature that radiographs digitized at pixel sizes smaller than 210μm are unsuitable for primary diagnosis, which is consistent with the above statement.

3. The equipment configuration available for study did not make it possible to resolve conclusively the relative influence on image quality of the display system and of the digitization process. Although both System 1 and System 2 were essentially "1K" monitors with access to "2K" digital data, it would therefore not be possible, from the data

presented here, to conclude that current "1K" monitors are necessarily intrinsically unsuitable for primary diagnosis.

4. It may be possible that better "1K" display systems, or "1K" display systems with a zoom facility providing a window into data of even higher spatial resolution, would yield more promising results, but this would require objective study. It is proposed that the effects on image perception of zoom, and many other psychophysical aspects of the use of display systems for diagnosis, remain important areas for further research, which should precede wider clinical use of such systems.

5. The data compression algorithm available on System 1 caused a significant reduction in diagnostic performance; no significant reduction in performance was demonstrated with the System 2 algorithm in the two image series studied.

6. It was observed that physical factors have a fundamental influence on diagnostic image quality, and some of these have been examined in detail. System 1 performed poorly in comparison with System 2, despite the fact that digitization to a pixel size of 100μm was possible with System 1 - double the resolution. The poor diagnostic performance was probably due to the interlacing of the monitors and to the perceived flicker; interlacing should be avoided in diagnostic display systems.

7. There is a pressing need for standardization of test patterns, test objects, and test data sets. There is also a need for manufacturers to provide more detailed technical data about their products, to permit an accurate comparison between different systems, to isolate features affecting performance, and to provide a baseline for quality assurance programmes.

8. With regard to the ROC methodology employed, it was possible to demonstrate that the use of paired data would not have altered the conclusions. It was also demonstrated that it is unwise to encourage observers to increase their efforts to use the points on ROC rating scales in a more uniform manner.

9. It is suggested that new methods of objective evaluation are badly needed. Requirements include much more realistic simulation of true radiological tasks, rather than just the conventional ROC issue of whether or not a pre-defined condition is present or absent, since radiologists are typically faced with a multiplicity of possible diagnoses, in any number of possible locations on any given image.

10. Image quality has an important bearing on the prospects and possible routes towards implementation of digital technology. An attempt has been made to consider the salient issues and to place them in context.

* * * * *

The goals of this study were to determine whether or not a significant loss of diagnostic image quality attended the use of commercially available digital radiological display systems, using genuine clinical case material. The hypothesis underlying such a study is that scope might exist for reducing image quality while preserving the diagnostic content of the image.

Objective evaluation of image quality is a complex and time-consuming procedure, even when the number of clinical conditions to be studied is small; in the course of the present investigation, using ROC methods, more than 13,000 individual observations were recorded. Obviously, clinical test data sets encompassing every conceivable radiological condition would not be possible, and ultimately one has to be content with well-validated samples of subtle case material.

These experiments show that the equipment studied was unsuitable for primary radiological diagnosis, and it is clear that many of parameters that influence diagnostic image quality have not yet been quantified.

It is likely that the objectives of total digital radiology will eventually be achieved, though at present there remain many unresolved issues. Work with "2K" display systems is encouraging, and it seems likely that the diagnostic performance of such systems will approach that of fast film-screen combinations. This cannot be assumed, however, and the need for careful, scientific assessment of PACS display systems prior to clinical introduction is inescapable.

REFERENCES

ACR-NEMA Standards Publication PS2: Data compression standards (1990). NEMA, Washington DC, USA

Adam A, Dawson P (1990) Clinical diagnosis from digital displays (letter). Clinical radiology **41** 143-4

Andrus W S, Dreyfuss J R, Jaffer F, Bird K T (1975) Interpretation of roentgenograms via interactive television. Radiology **116**:25-31

Aoki K, Nobechi T, Doi O, Fujimaki E, Mizuno T (1987) Pelvimetry. In Computed Radiography, Tateno Y, Iinuma T, Takano M (eds) Springer Verlag

Arenson R L, Chakraborty D P, Seshadri S B, Kundel H L (1990) The digital imaging workstation. Radiology **176**:303-315

Barnes G T, Sabbagh E A, Chakraborty D P, Nath P H, Luna R F, Sanders C, Fraser R G (1989) A comparison of dual-energy digital radiography and screen-film imaging in the detection of subtle interstitial pulmonary disease. Investigative radiology 24:585-591

Batnitzky S, Rosenthal S J, Siegel E L, Wetzel L H, Murphey M D, Cox G, McMillan J H, Templeton, A W, Dwyer, S J (1990) Teleradiology: an assessment. Radiology **177**:11-17

Beard D, Parrish D, Stevenson D (1990) A cost analysis of film image management and four PACS systems. Proceedings of the SPIE **1234:** 10-17

Berlin L (1980) Malpractice and radiologists. A J R **135:**587-91

Berlin L (1986) Malpractice and radiologists, update 1986: an 11.5 year perspective. American Journal of Roentgenology **147:**1291-1298

Blume H, Roehrig H, Browne M, Ji T L (1990) Comparison of the physical performance of high resolution CRT displays and films recorded by laser image printers and displayed on light boxes, and the need for a standard. Conference paper, SPIE Medical Imaging IV

Bramble J M, Cox G, Templeton A W, Murphey M D, Hensley K S, Anderson W H, Dwyer S J (1987) Data compression for digitally formatted radiographic examinations. Radiology **165(P)** 417 (conference paper)

Burgess A E (1988) Perception and detection of signals in medical images. NATO Advanced Study Institute (conference paper, available from author).

Capp, M P (1981) Radiological Imaging - 2000 AD. Radiology **138:**341-355

Capp M P, Roehrig H, Seeley G W, Fisher H D, Ovitt T W (1985) The digital radiology department of the future. Radiologic Clinics of North America 23, 2 349-354.

Capp M P, Nudelman S, Fisher H D, Ovitt T W, Pond G D, Frost M, Roehrig H, Seeger J, Oimette D (1981) Photoelectronic radiology department. SPIE 314 (Digital radiography) 2-8

Carboni N L, Tombaugh J W, Dillon R F (1989) The effect of simultaneous vs sequential viewing of digital images on comparative judgment performance. Proceedings of the SPIE **1091** 252-261

Carterette E C, Fiske R A, Huang, H K (1986) Receiver operating characteristic (R O C) evaluation of a digital viewing station for radiologists. S P I E proceedings, **626,** 441-446

Carmody D P, Nodine C F, Kundel H L (1980) Global and segmental search for lung nodules of different edge gradients. Invest. radiol. **15**:224-233

Chakraborty D P, Breatnach E S, Yester M V, Soto B, Barnes G T, Fraser R G (1986) Digital and conventional chest imaging: a modified ROC study of observer performance using simulated nodules. Radiology **158,** 35-39.

Chakraborty D P (1989), Maximum likelihood analysis of free-response receiver operating characteristic (FROC) data. Med. Phys **16** 561-8

Chakraborty D P, Winter L H (1990) Free-response methodology: alternate analysis and a new observer-performance experiment. Radiology **174**:873-81

Chan H, Vyborny C J, MacMahon H, Metz C E, Doi K, Sickles E A (1987). Digital mammography: ROC studies of the effects of pixel size and unsharp-mask filtering on the detection of microcalcifications. Investigative radiology **22**:581-589

Chan K, Taira R K, Huang H K (1990) Integration issues in PACS. In: PACS in Medicine: Proceedings of the NATO ASI. Huang H K, Ratib O, Bakker A R, Witte G (eds). In Press.

Cho P S, Ratib O, Ho B K, Morioka C A, Marx P, Huang H K, Shinagawa T, Sonobe T (1990) Personal digital image filing system. Proceedings of the SPIE **1234**:287-293

Cho P S, Huang H K, Tillisch J, Kangarloo H (1988) Clinical evaluation of a radiologic picture archiving and communication system for a coronary care unit. American Journal of Roentgenology **151**:823-7

Clark S, Kriz J S (1989) Displays for medical imaging. Information Display **5**:4-8

Cox G, Templeton A W, Dwyer S J (1986) Digital image management: networking, display and archiving. Radiologic Clinics of North America, **26**:37-54.

Cox G, Dwyer S J Templeton A W, (1986) Computer networks for image management in radiology. CRC critical reviews in diagnostic imaging. **25**:333-371

Cox G (1989) Interpretation of chest radiographs with a high resolution (2000 x 2000 x 12 bit) display. Radiology **173(P)**:402

Cox G, Cook L T, McMillan J H, Rosenthal S J, Dwyer S J (1990) Chest Radiography: comparison of high resolution digital displays with conventional and digital film. Radiology **176**: 771-76

Cohen M D, Long B, Cory D A, Broderick N J, Smith J A (1989) Digital imaging of the newborn chest. Clinical Radiology **40** 365-368

Craig, J O M C (1985) Diagnostic radiology without films. The Practitioner, **229**, 1011-15

Craig, J O M C (1988) The filmless department. British Journal of hospital medicine. **40**, 97-101

Craig J O M C, Highman J H, Glass H I (1986) Request for support for the evaluation of a digital image reporting console.

Dawood R M (1983) Radiography of 1252 patients in the Accident and Emergency Department at Guy's Hospital, London: an internal audit. Department of Radiology, Guy's Hospital.

Dawood R M, (1989) Radiology about to go digital: exciting new technology must be carefully evaluated (editorial) British Medical Journal, **299**: 340-1.

Dawood, R M (1990) Digital Radiology. (editorial) Clinical Radiology **42**, 6-11

Department of Health, UK (1989). Working for Patients: medical audit. Working paper 6, HMSO, London

Doi K, Giger M L, MacMahon H, Hoffman K R, Katsuragawa S, Yoshimura H, Nishikawa R N, F Yin, Metz C E, Asada N, Alperin N, Vyborny C J, Schmidt R A, Montner S, Chua K, Sanada S, Wu Y, Xu X, Carlin M (1990). Clinical radiology and computer-aided diagnosis: potential partners in medical diagnosis. Radiology **177(P)**:345

Fajardo L, Hillman B J, Hunter T B, Claypool H R, Westerman B R, Mockbee B (1987) Excretory urography using computed radiography. Radiology **162**:345-351

Fajardo L, Hillman B J, Pond G D, Carmody R F, Johnson J E, Ferrell W R (1989) Detection of pneumothorax: comparison of digital and conventional chest imaging. American Journal of Roentgenology **152**:475-80

Foley W D, Wilson C R, Keyes G S, DiBianca F A, Scanlon G T, Schleuter D, Lawson T L (1981) The effect of varying spatial resolution on the detectability of diffuse pulmonary nodules: assessment with digitized conventional radiographs. Radiology **141**:25-31

Fraser R G, Barnes G T, Hickey N M, Sabbagh E A, Sanders C, Niklason L T, Breatnach E, Chakraborty D (1987) Scanned projection digital radiography of the chest: a review of 5 years' experience. Alabama J. Med. Sci. **24**:258-66

Fraser R G, Sanders C, Barnes G T, MacMahon H, Giger M L, Doi K, Templeton A W, Cox G, Dwyer S J, Merritt C R B, Jones J (1989) Digital imaging of the chest. Radiology **171**, 297-307

Gale A G, Vernon J, Millar K, Worthington B S (1990) Reporting in a flash. Abstract. Proceedings of Radiology '90, British Institute of Radiology: p71

Giger M L, Doi K, MacMahon H, Metz C E, F Yin (1990) Pulmonary nodules: computer-aided detection in digital chest images. Radiographics. **10**:41-51

Glass H I (1990) Economic model of a whole hospital system. Proceedings of EuroPACS '90 (abstract).

Glass H I, Slark N A (1990) PACS and related research in the UK. In: PACS in Medicine: Proceedings of the NATO ASI. Huang H K, Ratib O, Bakker A R, Witte G (eds). In Press.

Goodman L R (1988a) Digital Radiography of the Chest: Promises and Problems. American Journal of Roentgenology **150**:1241-1252

Goodman L R, Foley W D, Wilson C R, Tikofsky R S, Hoffman R C (1988b) Pneumothorax and other lung diseases: effect of altered resolution and edge enhancement on diagnosis with digitized radiographs. Radiology **167**:83-88

Goodman L R, Foley W D, Wilson C R, Rimm A, Lawson T L (1986) Digital and conventional chest images: observer performance with film digital radiography system. Radiology **158**:27-33

Gordon M (1989) Dangers in the darkroom. The Radiographer **36**:114-5

Gray J E, Karsell P R, Becker G P, Gehring D G (1984) Total digital radiology: is it desirable or feasible? American Journal of Roentgenology **143**:1345-50

Gross G W, Ehrlich S M, Wang Y (1990) Diagnostic quality of portable abdominal radiographs in neonates with necrotizing enterocolitis. American Journal of Roentgenology **154**:779-83

Gur D (1989) New imaging and display techniques. Conference paper. Third Farwest Image Perception Conference, University of Arizona.

Hachiya J, Kato H (1982) Digital radiography using imaging plate. In Digital Radiography - Perspective for Clinical Applications. Akisada M (ed.) Life Science Centre, Tokyo.

Halpern E J, Newhouse J H, Amis E S, Levy H M, Lubetsky H W (1989) Evaluation of a quadtree-based compression algorithm with digitized urograms. Radiology **171**:259-263

Hansell D M (1990a) Digital chest radiography: Current status. Clinical Radiology **41**:229-231

Hansell D et al (1990b) Early experience with a prototype (2K) projectional digital chest unit: ROC study of simulated lung nodules. Poster presentation, BIR Annual Congress.

Hayrapetian A, Aberle D, Huang H K, Fiske R, Morioka C, Valentino D, Boechat M I, (1989a) A comparison of 2048-line digital display formats and conventional radiographs:

an ROC study. American Journal of Roentgenology **152** 1113-1118

Hayrapetian A, Morioka C, Huang H K, Fiske R, Aberle D, Boechat M I, Brown K, Hall T, Milos M, Tessler F (1989b) ROC Comparison of Analog Film to 512, 1024 and 2048 line Digital Displays. Proceedings of the SPIE **1092**:534-7

Ho B K, Chan K, Ishimitsu Y, Stewart B k, Lo S C, Huang H K (1987) High resolution image compression integrated into viewing station. Radiology **165(P)** 358 (conference paper).

Horii S C, Hill D G, Blume H R, Best D E, Thompson B G, Fuscoe C A (1990) ACR-NEMA standards activity: an update. SPIE Proceedings **1232** (In press)

Huang H K (1983) Digital Radiology at the University of California, Los Angeles: a feasibility study SPIE **418**:259-265

Huang H K (1987) Elements of digital radiology: a professional handbook and guide. Prentice Hall, New Jersey

Huang H K, Kangarloo H, Cho P, Taira R, Ho B, Chan K (1990) Planning a totally digital radiology department. AJR **154**:635-639

Huang H K, Mankovich N J, Cho P, Taira R, Stewart B K, Ho B (1987) PACS in Japan. AJR **148**:427-429

Huang H K, Cho P, Taira R, Ho B, Chan K (1990) PACS in Japan: 3 years later. AJR **154**:415-417

Huebener K H (1983) Scanned projection radiography of the chest versus standard film radiography: a comparison of 250 cases. Radiology **148**:363-368

Iinuma T (1974) Image processing in clinical medicine: consideration as a system. Medical Journal of Tokyo Women's College. **44** 152-9

Ishida M, Doi K, Loo L N, Metz C E, Lehr J L (1984) Digital image processing: effect on detectability of simulated low-contrast radiographic patterns. Radiology **150**:569-75

Ishigaki T, Sakuma S, Ikeda M, Itoh Y, Suzuki M, Iwai S (1990) Clinical evaluation of irreversible image compression: analysis of chest imaging with computed radiography. Radiology **175**:739-43

Johansson T, Nilsson J, Holmer N G, Hertz C H (1990) Ink jet printing of digital radiographs. Acta radiologica. **31** 108-110

Johnston R F, Pickett S C, Barker D L (1990) Autoradiography using storage phosphor technology. Electrophoresis **11**:355-60

Judy P F, Swensson R G, Szulc M (1981) Lesion detection and signal to noise ratio in CT images. Med. Phys **8**:12-23

Judy P F, Swensson R G, Chopra P (1989) Grading stenotic lesions: a perceptual study. Abstract. Radiology **173(P)** 225

Kangarloo H, Boechat M I, Barbaric Z, Taira R K, Cho P S, Mankovich N J et al (1988) Two year clinical experience with a computed radiography system. American Journal of Roentgenology **151:** 606-608.

Kastan D J, Ackerman L V, Feczko P J (1987) Digital gastrointestinal imaging: the effect of pixel size on detection of subtle mucosal abnormalities. Radiology **162:** 853-6

Kazan B (1989) New electronic displays and their possibilities for radiology. Investigative Radiology. **24** 817-9

Ker M, Seeley G W, Stempski M, Patton D (1988) A protocol for verifying truth of diagnosis. Investigative radiology, **23,** 485-487

Kimme-Smith C, Gold R H, Bassett, Gormley L, Morioka C (1989a) Diagnosis of Breast Calcifications: Comparison of Contact, Magnified, and television Enhanced Images. AJR **153**:963-967.

Kimme-Smith C, Bassett, Gold R H, Gormley L, (1989b) Digital mammography: Comparison of two methods. AJR **153**:963-967.

Kogutt, S, Jones J P, Perkins, D (1988) Low dose digital computed radiography in pediatric chest imaging. American Journal of Roentgenology **151,** 775-779.

Kogutt M S (1987a) Use of a computed radiography system for low-dose digital radiography of infants and children. Radiology **165(P):** 415-6

Kogutt M S (1987b) Computed radiographic imaging: use in low-dose leg length radiography. American Journal of Roentgenology **148,** 1205-1206.

Kosslyn S M (1989) The psychology of visual displays. Investigative Radiology. **24**, 417-9

Krupinski E A, Nodine C F, Kundel H L (1989) Visual feedback enhances the recognition of missed lung nodules in chest X-ray images. Poster presentation, Third Farwest Image Perception Conference, University of Arizona.

Kundel H L (1979) Images, Image Quality and Observer performance. Radiology **132**:265-271

Kundel H L, Seshadri S B, Arenson R L (1990) Clinical evaluation of PACS: modeling diagnostic value. Proceedings of the SPIE **1234:** 214-7

Kundel H L (1986a) Visual perception and image display terminals. Radiologic Clinics of North America. **24**, 69-78.

Kundel H L (1986b) Digital projection radiography of the chest. Radiology **158**:274-6

Kundel H L, Mezrich J L, Brikman I, Siegel R, Miller W T, Epstein D M, Gefter W B, Arenson R L, Seshadri S B, Khalsa S (1987) Digital chest imaging: comparison of two film image digitizers with a classification task. Radiology **165**:747-52

Lams, P M, Cocklin, M L (1986) Spatial resolution requirements for digital chest radiographs: an ROC study of observer performance in selected cases. Radiology, **158,** 11-19.

Levin K, Fielding R F (1989) Methods to prefetch comparison images in an image management and communication system. Proceedings of the SPIE **1234**:270-274

Lo S K, (1986) Compression of Radiological Images with 512, 1024 and 2048 Matrices. Radiology **161** 519-525

MacMahon H, Vyborny C J, Sabeti V, Solomon S L (1986) Digital radiography of subtle pulmonary abnormalities: an ROC study of the effect of pixel size on observer performance. Radiology **158,** 21-26.

MacMahon H, Metz C E, Doi K, Kim T, Giger M, Chan H (1988) Digital chest radiography: effect on diagnostic accuracy of hard copy, conventional video, and reversed gray scale video display formats. Radiology **168**:669-673

MacMahon H, Doi K, Sanada S, Montner S, Giger M L, Metz C E, Nakamori N, F Yin, Xu X, Yonekawa H, Takeuchi H (1991). Data compression: effect on diagnostic accuracy in digital chest radiography. Radiology **178**:175-9

McMillan J H, Huang H K, Bramble J M, Siegel E L (1989) Digital Radiography. Investigative radiology **24**:735-41.

MDIS Acquisition Document (1990) US Army Engineer Division, Huntsville, Alabama, USA

Metz, C E (1978) Basic principles of R O C analysis. Seminars in nuclear medicine. **8**:283-298

Metz, C E (1986) R O C methodology in radiologic imaging. Investigative radiology, **21:**720-731

Metz, C E (1987) ROCFIT software package, University of Chicago.

Metz, C E (1989) Some practical issues of experimental design and data analysis in radiological R O C studies. Investigative radiology **24** 234-235

Miller E R, McCurry E M (1969) Immediate hospital-wide access to X-ray film images. Radiology **92** :225-230

Mosser H, Mandl A, Urban M, Hradil H, Hruby W (1990) The Vienna SMZO project. In: PACS in Medicine: Proceedings of the NATO ASI. Huang H K, Ratib O, Bakker A R, Witte G (eds). In Press.

Mun S K, Akisada M (1989) Japanese PACS development and implementation activities. Conference paper, IMAC '89, Washington DC. Available from authors.

Murphey M D (1989) Digital skeletal radiography: spatial resolution requirements for detection of subperiosteal resorption. American Journal of Roentgenology **152**:541-546

Murphey M D, Bramble J M, Cook L T, Martin N L, Dwyer S J (1990) Nondisplaced fractures: spatial resolution requirements for detection with digital skeletal imaging. Radiology **174**:865-870

Newell J D, Seeley G, Hagaman R M et al (1988) Computed radiographic evaluation of simulated pulmonary nodules: preliminary results. Investigative radiology **23**:267-70

Oestmann J W, Kushner D C, Bourguin P M, Llewellyn H J, Mockbee B W, Greene R (1988a) Subtle Lung Cancers: Impact of Edge Enhancement and Grey Scale Reversal on Detection with Digitized Chest Radiographs. Radiology **167**:657-658

Oestmann J W, Kopans D, Hall D, McCarthy K, Rubens J, Greene R (1988b) A comparison of digitized storage phosphors and conventional mammography in the detection of subtle microcalcifications. Investigative radiology **23**:725-728

Oestmann J W, Greene R, Rhea J T, Rosenthal H, Koenker R M, Tillotson C L et al (1989a) Single exposure dual energy digital radiography in detection of pulmonary nodules and calcifications. Investigative Radiology. **24**, 517-21

Oestmann J W, Rubens J R, Bourguin P M, Rhea J T, Llewellyn H J, Greene R, (1989b) Impact of post processing on detection of simulated pulmonary nodules with digital radiography. Investigative Radiology. **24**, 467-471

Oestmann J W, Greene R, Rubens J R, Pile-Spellman E, Hall D, Robertson C, Llewellyn H J et al (1989c) High frequency edge enhancement in the detection of fine pulmonary lines. Investigative Radiology. **24**, 643-646

Okabe T (1990) PACS and related research in Japan. In: PACS in Medicine: Proceedings of the NATO ASI. Huang H K, Ratib O, Bakker A R, Witte G (eds). In Press.

Pountain D (February 1989) Digital paper. BYTE magazine, 274-280

Ratib O, Ligier Y, Appel R, Scherrer J R (1990) Papyrus: a portable image file format. In: PACS in Medicine: Proceedings of the NATO ASI. Huang H K, Ratib O, Bakker A R, Witte G (eds). In Press.

Revesz R and Kundel H L (1973) Transmission of radiographic images via closed circuit television and microwave television techniques. Investigative radiology **8**:392-95

Roehrig H, Dallas W, Ji T L, Lamoreaux R D, Oikawa D, Vercillo R, Yocky D (1989) Physical evaluation of CRTs for use in digital radiography. Proceedings of the SPIE **1091** 262-278

Roehrig H, Blume H, Ji T L, Browne M (1990) Performance tests and quality control of cathode ray tube displays. Journal of Digital Imaging **3:** 134-45

Sakuma, H, Takeda, K, Hirano, T, Okuda, Y, Matsumura, K, Yamaguchi, N et al (1988) Plain chest radiograph with computed radiography: improved sensitivity for the detection of coronary artery calcification. American Journal of Roentgenology **151,** 27-30.

Schaefer C M, Greene R, Oestmann J W, Hall D, Llewellyn H J McCarthy K A, Pile-Spellman E, Rubens J R (1990) Digital storage phosphor imaging versus conventional film radiography in CT-documented chest disease. Radiology **174**:207-210

Seeley G W, Roehrig H, Nudelman S, Frost M, Capp M P (1978) Psychophysical evaluation of video systems compared to film. Appl. Photogr. Eng. **4**:160-165.

Seeley G W, Newell J D (1986) The use of psychophysical principles in the design of a total digital radiology department. Radiologic Clinics of North America **23**:341-348

Seeley G W, Fisher H D, Stempski M O, Borgstrom M, Bjelland J, Capp M P (1987) Total digital radiology department: spatial resolution requirements. American Journal of Roentgenology **148**:421-426.

Seeley G W, Ovitt T, Capp M P (1984) The Total Digital Radiology Department: An alternative view. AJR **144**: 421-422.

Sheline M E, Brikman I, Epstein D M, Mezrich J L, Kundel H L, Arenson R L (1989) The diagnosis of pulmonary nodules: comparison between standard and inverse digitized images and conventional chest radiographs. American Journal of Roentgenology **152**:261-263

Shultze Kool L, Huang H K (1990) Developments in image acquisition systems. In: PACS in Medicine: Proceedings of the NATO ASI. Huang H K, Ratib O, Bakker A R, Witte G (eds). In Press.

Smathers R. Mammographic Microcalcification: detection with Xerography, Screen film and Digitized Film Display. Radiology **159**:673-677

Slasky B S, Gur D, Good W F, Costa-Greco M A, Harris K M, Cooperstein L A, Rockette H E (1990). Receiver operating characteristic analysis of chest image interpretation with conventional, laser-printed, and high resolution workstation images. Radiology **174**:775-80

Sonoda M, Takano M, Miyahara J, Katoh H (1983) Computed radiography utilizing scanning laser stimulated luminescence. Radiology **148**:833-8

Spackman, T J, Bensman, K W (1987) Development of a useful picture archiving and communications system. American Journal of Roentgenology **148**:1025-1026.

Spring D B, Tennenhouse D J (1986) Radiology malpractice lawsuits: California jury verdicts. Radiology **159**:811-14

Steckel R J (1972) Daily rounds in a large teaching hospital using high resolution closed circuit television Radiology **105**:319-321

Stockburger W T, King W E (1990). PACS: a financial analysis for economic viability. Applied radiology, January 1990

Swets J A (1979) ROC Analysis Applied to the Evaluation of Medical Imaging Techniques. Invest. Radiology **14**:2 109-121

Swets, J A and Pickett, R M (1986). Evaluation of diagnostic systems: methods from signal detection theory. Academic Press, London

Taira, R K, Mankovich, N J, Boechat, M I, Kangarloo H, Huang, H K (1988) Design and implementation of a picture archiving and communication system for pediatric radiology. American Journal of Roentgenology **150**:1117-1121.

Takano M (1982) New computed radiography: Fuji intelligent diagnostic system. Journal of medical imaging **2**:202-7

Takahashi K, Kohda K, Miyahara J, Kanemitsu Y, Amitani K, Shionya S (1984). Mechanism of photostimulated luminescence in $BaFX;Eu^{2+}$ (X=Cl, Br). Phosphors. J. Lumi. **31**:266

Tateno Y (1984). Clinical application of FCR. Journal of medical imaging. Vol 4 Suppl1 7-15

Tateno Y, Iinuma T, Takano M (eds) (1987). Computed Radiography, Springer Verlag

Templeton, A W, Cox, G, Dwyer, S J, (1988) Digital image management networks: current status. Radiology. **169,** 193-199.

Templeton, A W, Dwyer, S J, Cox, G, Lee, K R, Johnson, J A, Martin, N L et al (1987) A digital radiology imaging system: description and clinical evaluation. American Journal of Roentgenology **149**:847-851.

Ukovich W, Dalla Palma L, Stacul F, Zernich R C, Giribona P (1990) Practical experience with PACS: benefits and pitfalls. Proceedings of the SPIE **1234**:34-43

van Gennip E (1990) A cost-benefit study of PACS: a model-wise approach. Conference paper, EUROPACS '90, Trieste.

van Poppel B M, Lodder H, Bakker A (1990) Implications of image management strategies for standardization. Proceedings of ISCAMI '90. (abstract)

Vlasbloem H, Schultze Kool L J (1988) AMBER: a scanning multiple-beam equalization system for chest radiography. Radiology **169** 29-34

Wald A. Statistical Decision Functions. New York: Wiley 1950

Wendler T, Wein B (1990) A rule-base for the co-operative selection and display of images at PACS workstations. Proceedings of EuroPACS '90. (abstract)

Wilenzick R M, Merritt CR, Balter S (1987) Megavoltage portal films using computed radiographic imaging with photostimulable phosphors. Medical Physics **14** 389-92

Yonekawa H, Takeuchi H (1988) Medical image compression with adaptive block cosine transform coding. Radiology **165(P)** 358

APPENDIX 1: LIST OF ABBREVIATIONS

μm	micron
1K 2K etc	digitization matrix of 1000 x 1000, 2000 x 2000
ACR	American College of Radiology
CR	computed radiography
CRT	cathode ray tube
CLAHE	contrast limited adaptive histogram equalization
CT	computerized tomgraphy
Db	decibel
DPCM	differential pulse code modulation
DCT	discrete cosine transform
FDA	Food and Drug Administration (USA)
FRCR	Fellow, Royal College of Radiologists
ft-Lambert	foot-Lambert
HIS	hospital information system
ICU	intensive care unit
IMAC	image management and communication
IMS	image management system
lp/mm	line pairs per millimeter
MRI	magnetic resonance imaging
NEMA	National Electrical Manufacturers' Association
PACS	picture archiving and communication systems
PCP	Pneumocystis carinii pneumonia
RIS	radiology information system
ROC	receiver operating characteristic
UCLA	University of California at Los Angeles
VDU	visual display unit

APPENDIX 2: ROC ANALYSIS: A BRIEF EXPLANATION OF THE ROC CURVE

Consider the distribution of a biological variable in a healthy population and in disease. The precise nature of the variable hardly matters; it could for example be haematological - such as haemoglobin measurements in normal patients and patients with polycythaemia; or biochemical, or even radiological; in all of these situations there is a normal and abnormal range, and the two overlap (Figure i and ii).

In discriminating between normal and abnormal cases on the basis of a test result, the position of the decision criterion is critical. A lax definition of abnormality will create many false positives. A stricter definition will permit many false negatives.

It is not really necessary to consider true negatives and false negatives separately; their relationship is embraced in the relationship between true and false positives (Figure iii).

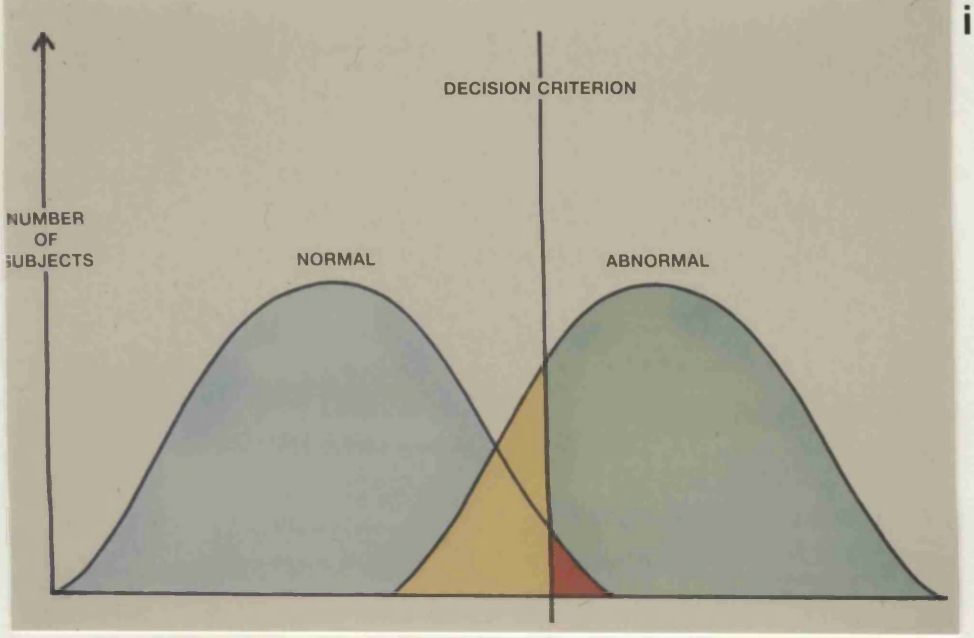
ROC analysis explores the relationship between true positive rate and false positive rate as the decision criterion varies.

The ROC curve is a graph of true positive rate against false positive rate. Figure iv shows a curve for a test that fails to discriminate between normal and abnormal - in fact it is a straight line.

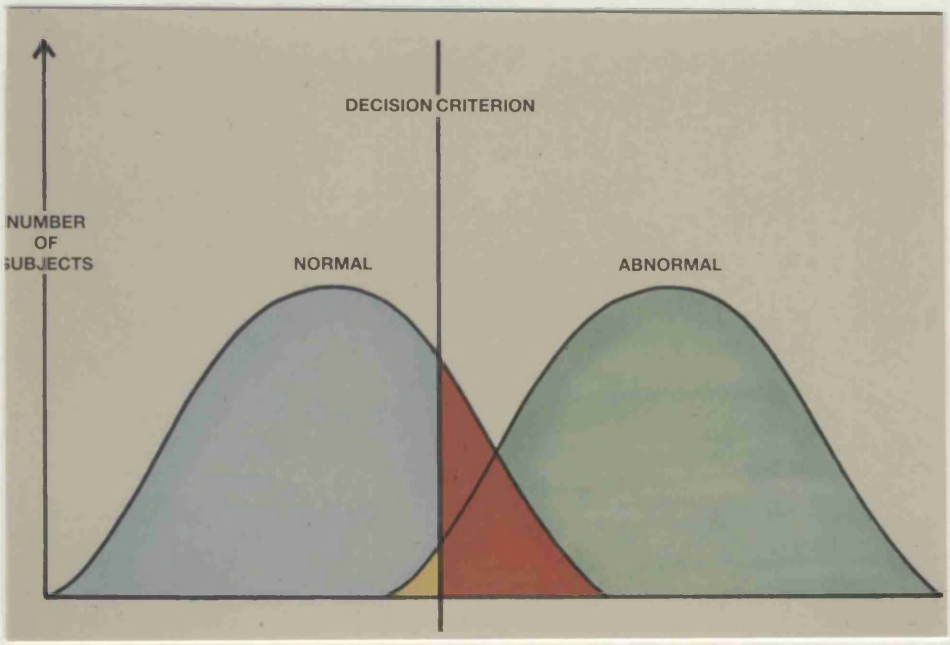
Figure v shows more typical curves for three more tests - the better the test, the more closely the curve approaches the upper left corner. The vertical line allows comparison of the true positive rate for these three tests at false positive value of 15% - approximately 80%, 60% and 35%.

The curves can be used to compare the performance of different types of display systems with each other and with film.

i



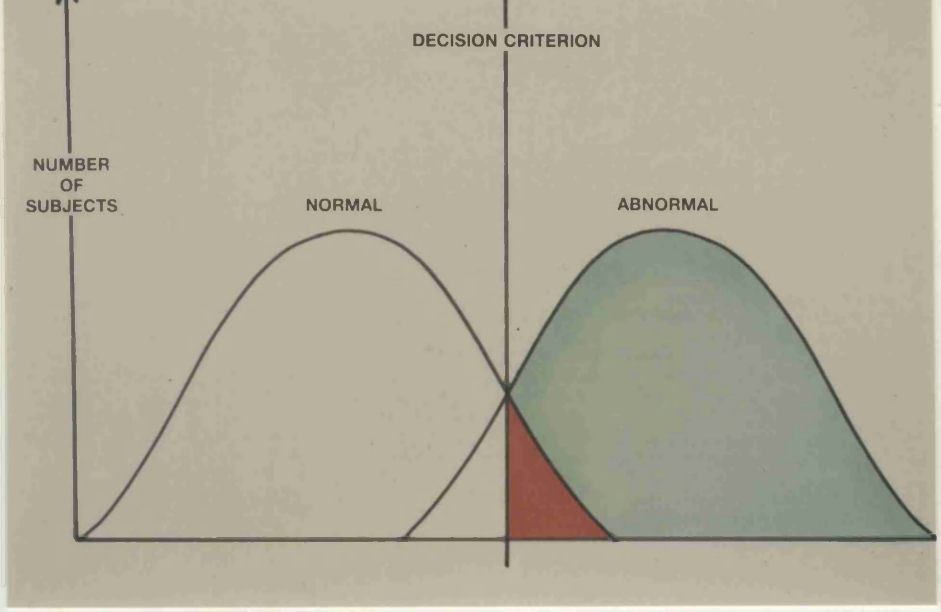
ii



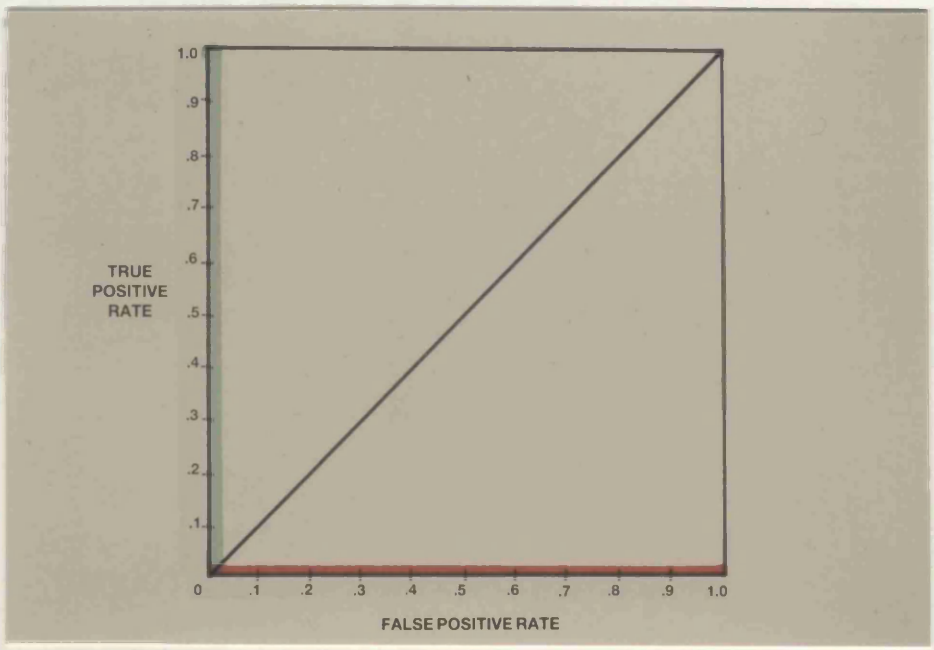
KEY

	TRUE POSITIVE (abnormal cases, correctly diagnosed)
	FALSE POSITIVE (normal cases, misdiagnosed)
	TRUE NEGATIVE (normal cases, correctly recognised)
	FALSE NEGATIVE (abnormal cases, not recognised)

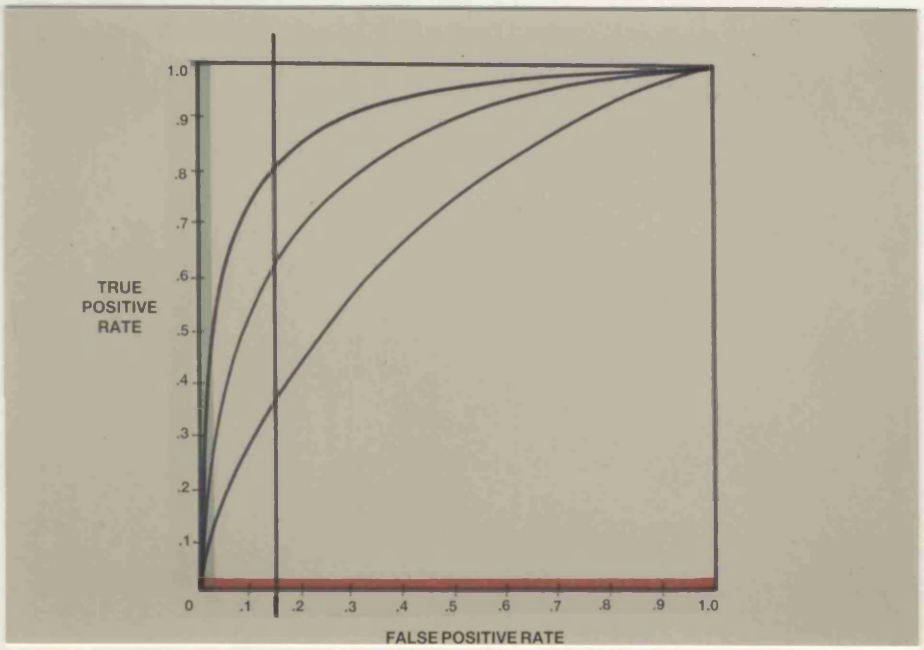
iii


DECISION CRITERION
NUMBER
OF
SUBJECTS
NORMAL
ABNORMAL

iv


TRUE
POSITIVE
RATE
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1.0
FALSE POSITIVE RATE

v


TRUE
POSITIVE
RATE
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1.0
FALSE POSITIVE RATE

APPENDIX 3: PUBLICATIONS & PRESENTATIONS RELATING TO THE ST MARY'S DIGITAL PROJECT

1. Diagnostic radiology without films. J O M C Craig (1985) The Practitioner, 229, 1011-15

2. The filmless department. J O M C Craig (1988) British journal of hospital medicine. **40**, 97-101

3. Clinical diagnosis from digital displays: the St Mary's evaluation project. R M Dawood. Invited oral presentation at the Annual Scientific Meeting of the Royal College of Radiologists, Exeter, September 1988.

4. Clinical diagnosis from digital displays: preliminary findings of the St Mary's evaluation project. R M Dawood, J O M C Craig, J H Highman, J Wadsworth, H I Glass, A Todd-Pokropek, D A Cunningham, J M Stevens, A Al-Kutoubi, R W Kerslake, A H Choudhri, C J Barber, M E Crofton, A W Porter. Clinical Radiology, 1989. **40**: 369-373

5. Monitor quality and its importance in a Picture Archiving and Communications System (PACS) display workstation. Dawood R M, Todd-Pokropek A, Craig J O M C, Highman J H, Glass H I. Oral presentation, 74th Scientific Assembly of the Radiological Society of North America, November 1988. Radiology, 1989: **169(P)**, 358

6. Evaluation of VDU displays in radiological diagnosis: a preliminary comparison of two systems. R M Dawood, J O M C Craig, J H Highman, J Wadsworth, H I Glass, A Todd-Pokropek, D A Cunningham, J M Stevens, A Al-Kutoubi, R W Kerslake, C J Barber, M E Crofton, A W Porter. Poster presentation. Medical Imaging III, SPIE (Society of Photo-Optical Instrumentation Engineers) Annual Meeting 1989, Newport Beach, USA

7. Evaluation of VDU displays in radiological diagnosis: a preliminary comparison of two systems. R M Dawood, J O M C Craig, J H Highman, J Wadsworth, H I Glass, A Todd-Pokropek, D A Cunningham, J M Stevens, A Al-Kutoubi, R W Kerslake, C J Barber, M E Crofton, A W Porter. Proceedings of the SPIE: **1091**: 638-371.

8. The evaluation of digitizer and video display systems in radiology: tests and test data sets. A. Todd-Pokropek, R. M. Dawood, J.O.M.C. Craig, H. Highman, H.I. Glass. Oral presentation, EUROPACS 89, Brussels.

9. Radiographic practice in digital research. Oral presentation, Annual Meeting of the Society of Radiographers, Eastbourne, 1989.

10. Radiological diagnosis from digital displays: an R O C study of images from patients with <u>Pneumocystis carinii</u> pneumonia. R M Dawood, J O M C Craig, J H

Highman, J Wadsworth, A Todd-Pokropek, A W Porter, H I Glass, M McCarty, D A Cunningham, J M Stevens, A Al-Kutoubi, C J Barber, M E Crofton, P C Rowlands, T Tran. Poster presentation, IMAC '89, Washington DC, June 1989.

11. The turbulent road to a filmless hospital. J O M C Craig, R M Dawood. Invited presentation, CAR (Computer Assisted Radiology) '89, Berlin, June 1989

12. The turbulent road to a filmless hospital. J O M C Craig, R M Dawood. In CAR '89: Proceedings of the international symposium. H U Lemke, M L Rhodes, C Jaffe, R Felix (Eds.) Springer-Verlag 1989

13. Radiographic implications of digital radiology. A W Porter. Oral presentation, ISSRT, Paris, July 1989

14. Radiology about to go digital: exciting new technology must be carefully evaluated. R M Dawood. Editorial, British Medical Journal, 1989. **299**: 340-1.

15. Radiology about to go digital. R M Dawood. Letter, British Medical Journal, 1989 **299**:854

16. Image quality on PACS workstations: results with clinical diagnosis. R M Dawood, J O M C Craig, J H Highman, A Todd-Pokropek, A W Porter, J Wadsworth, D A Cunningham, J M Stevens, A Al-Kutoubi, C J Barber, M E

Crofton, M McCarty, P C Rowlands, T Tran. Oral presentation at the Annual Scientific Meeting of the Royal College of Radiologists, Liverpool, September 1989.

17. Clinical evaluation of data compression and its influence on displayed image quality in PACS. R M Dawood, J O M C Craig, J H Highman, H I Glass, A Todd-Pokropek, J Wadsworth, A W Porter, D A Cunningham, J M Stevens, A Al-Kutoubi, C J Barber, M E Crofton, P C Rowlands, T Tran. Oral presentation, 75th Scientific Assembly of the Radiological Society of North America, November 1989. Radiology **173 (P)** 258

18. Does pooling data change the validity of ROC assessments? A Todd-Pokropek, R M Dawood. Oral presentation, 75th Scientific Assembly of the Radiological Society of North America, November 1989. Radiology **173 (P)** 225

19. Clinical evaluation of diagnostic image quality from PACS workstations. J H Highman, R M Dawood, J O M C Craig, H I Glass, A Todd-Pokropek, J Wadsworth, A W Porter, D A Cunningham, J M Stevens, A Al-Kutoubi, C J Barber, M E Crofton, P C Rowlands, T Tran. Oral presentation, 75th Scientific Assembly of the Radiological Society of North America, November 1989. Radiology **173 (P)** 429

20. Perception of noise on a PACS display, and the influence on signal to noise ratio of a film digitizer. R M Dawood, A Todd-Pokropek, J O M C Craig, J H Highman, A W Porter. Proceedings of the SPIE, 1990: **1234**:947-51

21. Perception of noise on a PACS display, and the influence on signal to noise ratio of a film digitizer. R M Dawood, A Todd-Pokropek, J O M C Craig, J H Highman, A W Porter. Poster presentation. Medical Imaging IV, SPIE (Society of Photo-Optical Instrumentation Engineers) Annual Meeting 1990, Newport Beach, USA

22. Optimal use of rating scales in ROC analysis. R M Dawood, A Todd-Pokropek, J O M C Craig, J H Highman, A W Porter. Poster presentation. Medical Imaging IV, SPIE (Society of Photo-Optical Instrumentation Engineers) Annual Meeting 1990, Newport Beach, USA

23. Optimal use of rating scales in ROC analysis. R M Dawood, A Todd-Pokropek, J O M C Craig, J H Highman, A W Porter. Proceedings of the SPIE, 1990: **1234**:952-56

24. Digital Radiology: current problems. Invited presentation, Institute of Physical Sciences In Medicine Joint Scientific Meeting, Newcastle, April 1990.

25. Digital Radiology: current problems. R M Dawood. Physics in Medicine & Biology, 1990. In Press.

26. Quality assurance in PACS. R M Dawood, J O M C Craig, J H Highman, A Todd-Pokropek, A W Porter. Conference paper, Annual Scientific Meeting of the British Institute of Radiology, Harrogate, June 1990.

27. PACS in Britain. Invited presentation, NATO Advanced Study Institute, Evian, October 1990.

28. Digital radiology. R M Dawood. (1990) Editorial, Clinical Radiology **42**, 6-11

29. Criteria for selection of clinical images for assessment of image quality in PACS. R M Dawood, A Todd-Podropek, J H Highman, A W Porter, J O M C Craig. Oral presentation, EuroPACS '90, Trieste.

30. Medico-legal implications of digital radiology. J O M C Craig, R M Dawood. Oral presentation, ISCAMI, Crete, May 1990.

31. Clinical diagnosis from digital displays. R M Dawood, (1990). Letters, Clinical Radiology, 41:143-4

32. Are 1K display systems suitable for primary diagnosis from radiographic images in PACS? R M Dawood, J O M C Craig, J H Highman, A Todd-Pokropek, J Wadsworth, A W

Porter, D A Cunningham, J M Stevens, A Al-Kutoubi, C J Barber, M E Crofton, P C Rowlands. Poster presentation, 76th Scientific Assembly of the Radiological Society of North America, November 1990 Radiology **177 (P)** 135

33. Test data sets for the evaluation of systems in PACS. A Todd-Pokropek and R M Dawood, Poster presentation, 76th Scientific Assembly of the Radiological Society of North America, November 1990 Radiology **177 (P)** 247

APPENDIX 4:PUBLICATIONS IN SUPPORT

1. Clinical diagnosis from digital displays: preliminary findings of the St Mary's evaluation project. R M Dawood, J O M C Craig, J H Highman, J Wadsworth, H I Glass, A Todd-Pokropek, D A Cunningham, J M Stevens, A Al-Kutoubi, R W Kerslake, A H Choudhri, C J Barber, M E Crofton, A W Porter. Clinical Radiology, 1989. **40**: 369-373

2. Radiology about to go digital: exciting new technology must be carefully evaluated. R M Dawood. Editorial, British Medical Journal, 1989. **299**: 340-1. (5 August 1989)

3. Digital Radiology. Dawood, R M (1990) Editorial, Clinical Radiology **42**, 6-11

4. Evaluation of VDU displays in radiological diagnosis: a preliminary comparison of two systems. R M Dawood, J O M C Craig, J H Highman, J Wadsworth, H I Glass, A Todd-Pokropek, D A Cunningham, J M Stevens, A Al-Kutoubi, R W Kerslake, C J Barber, M E Crofton, A W Porter. Proceedings of the SPIE, 1989: **1091**: 638-371.

5. Perception of noise on a PACS display, and the influence on signal to noise ratio of a film digitizer. R M Dawood, A Todd-Pokropek, J O M C Craig, J H Highman, A W Porter. Proceedings of the SPIE, 1990: **1234**:947-51

6. Optimal use of rating scales in ROC analysis. R M Dawood, A Todd-Pokropek, J O M C Craig, J H Highman, A W Porter. Proceedings of the SPIE, 1990: **1234**:952-56

Clinical Radiology (1989) **40**, 369–373

Clinical Diagnosis from Digital Displays: Preliminary Findings of the St Mary's Evaluation Project

R. M. DAWOOD, J. O. M. C. CRAIG, J. H. HIGHMAN, J. WADSWORTH*, H. I. GLASS†, A. TODD-POKROPEK‡, D. A. CUNNINGHAM, J. M. STEVENS, A. AL-KUTOUBI, R. W. KERSLAKE, A. H. CHOUDHRI, C. J. BARBER, M. C. CROFTON and A. W. PORTER

*Department of Radiology, St Mary's Hospital, London, *Academic, Department of Community Medicine, St Mary's Hospital Medical School, †North West Thames Regional Health Authority and ‡Department of Medical Physics, University College London*

Image quality is a fundamental issue in the introduction of picture archiving and communications systems (PACS), and one that has hitherto been eclipsed by other aspects of the considerable technological challenge facing scientists and manufacturers involved in its development. We conducted a formal evaluation of clinical radiological diagnosis from a commercially available PACS viewing station, using subperiosteal resorbtion in renal osteodystrophy as the test pathological diagnosis, with receiver operating characteristic (ROC) analysis of the results. We conclude that the displayed, digitised images were inferior to film using the apparatus tested and believe that careful, objective clinical evaluation of such systems is of paramount importance.

The proposed redevelopment of St Mary's Hospital coincided with a period of rapid progress, awareness and interest in digital imaging technology, and seemed to provide an ideal opportunity to implement the PACS ideal in its entirety (Craig 1985, 1988).

Although the pace of progress in PACS technology did not enable the realisation of this objective in the time available, the new department has been planned to take advantage of likely future developments, and included a PACS workstation and archive for evaluation.

The potential benefits of PACS have been considered in detail elsewhere (e.g. Capp *et al.*, 1985; Huang, 1987). All radiologists are familiar with the frustrations associated with working with film, but it is worth bearing in mind that the potential benefits are not just logistic and administrative: digital radiography offers the promise of lower patient radiation dosage (reduced by up to 98% in some examinations) (Kogutt, 1987; Kogutt *et al.*, 1988) and opens up the prospect of being able to interrogate and enhance images for improved diagnosis.

There are many reasons why technological progress has been relatively slow; initially, many of the problems were seriously underestimated, and there is clearly more to a PACS network than merely joining up items of equipment with lines on a diagram (Cox *et al.*, 1986; Spackman & Bensman, 1987; Templeton *et al.*, 1988). While filmless radiography, with reusable photostimulable phosphor cassettes, has been developed to a point of high sophistication (Kangarloo *et al.*, 1988) there are still enormous problems that relate to the huge amounts of digital data involved: data storage, traffic and retrieval.

Worldwide, most of the centres that are attempting to introduce and develop PACS systems are run by physicists, engineers and computer scientists. At St Mary's we chose to focus upon one single, crucial element of the digital chain, the digital display. As diagnostic radiologists, we set out to evaluate a commercially available system from the point of view of its clinical radiological diagnostic value. The most important question, as far as diagnostic radiologists and their patients are concerned, is one that is often neglected: will we be able to make the correct diagnosis?

Correspondence and reprint requests to: Dr R. M. Dawood, Department of Radiology, St Mary's Hospital, London W2 1NY.

METHODS

The equipment installed at St Mary's consisted of two laser film digitisers (scanning spot size approximately 100 microns) which are in fact identical, but configured to scan at different matrix sizes—nominally 1 K and 2 K, an optical disc archive, with a single disc drive (1 GB per side) and a 160 MB magnetic disc buffer, an image management system, to link patient data with the images themselves, and a display terminal consisting of a four-monitor viewing station (Fig. 1) with 160 MB magnetic disc for local storage.

1024 line monitors were used, with a band width corresponding to a 48 mHz pixel clock rate and 8-bit video coding.

Scan matrix sizes vary according to film size. Small film sizes were used in our evaluation study and Table 1 shows the matrix sizes that resulted. All references to '1 K' and '2 K' in this paper in fact exaggerate the matrix size in each case by a small degree.

In order to test a system to its limits, it is necessary to choose a pathological condition to study that is subtle, and difficult to diagnose even on film. We chose subperiosteal resorbtion in renal osteodystrophy, one of several conditions used by other workers who have studied the effects of digitisation upon the diagnostic process. Previous work has, however, largely centred upon comparisons between film and digitised hard copy, or on images prepared using phantoms (e.g. Chakraborty *et al.*, 1986; MacMahon *et al.*, 1986; Carterette *et al.*, 1986; Seeley *et al.*, 1987; Sakuma *et al.*, 1988).

We used a series of 40 hand radiographs. Half were from patients who had no known renal disease, and who had attended the X-ray department for a variety of unrelated conditions; the other half were from patients with proven chronic renal failure in whom a diagnosis of subperiosteal resorbtion had been evident radiologically on at least two occasions and in whom the diagnosis was also visible to the radiologist administering the tests (RMD); the cases were also discussed with radiologists not participating in the study. Patients with obvious metaphyseal changes or other ancillary features were excluded. Some of the cases selected were subtle, and intentionally so.

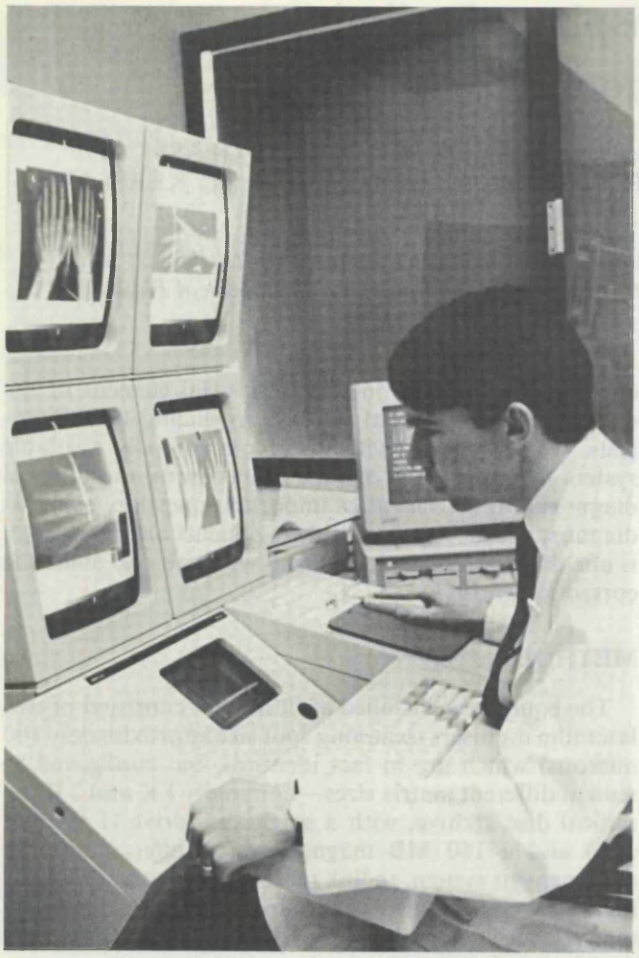


Fig. 1 – Viewing station in use during the study.

Table 1 – (*a*) and (*b*). Digitisation matrix sizes for film sizes used.

	'1 K'	'2 K'
(a) Digitisation parameters for film size 18 × 24 cm (8" × 10")		
Area scanned	17·5 × 22·8 cm	
Pixels read	2000 × 2600	
Pixels stored	750 × 975	1500 × 1950
Pixel size	0·2333 mm	0·117 mm
(b) Digitisation parameters for film size 24 × 30 cm (10" × 12")		
Area scanned	24·7 × 28·8 cm	
Pixels read	1800 × 2270	
Pixels stored	752 × 908	1504 × 1816
Pixel size	0·328 mm	0·164 mm

The original films were digitised as '1 K' and '2 K'.

Besides investigating the effect of matrix size on diagnosis we also had an opportunity to examine the effect of data compression, since the ability to compress data has important economic and practical consequences. A proprietary cosine transform algorithm was built into our unit, enabling typical compression ratios of 15:1. Having seen the '1 K' compressed images, we decided that these were not of adequate quality to merit study, but decided to include a study of '2 K' compression in our evaluation.

There were effectively four different image formats under study; original film, '1 K' and '2 K' uncompressed, and '2 K' compressed.

The images were presented to a team of eight observers.

Before evaluation of the system could be undertaken, it was essential for the observers to become familiar with the equipment, and the techniques of displaying and manipulating the images, and recording observations in a standard way suitable for ROC analysis. This process was monitored by using a test series of images—in fact a short

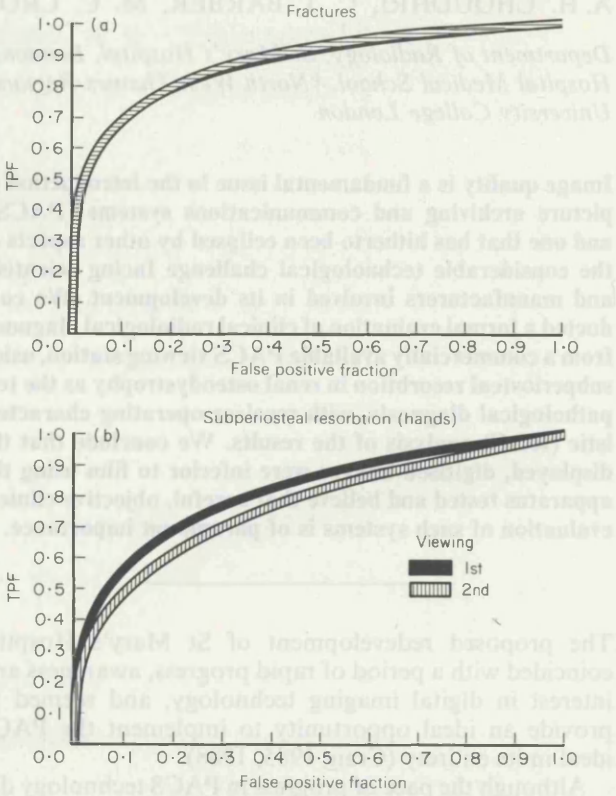


Fig. 2 – (*a*) ROC curves from casualty cases. (*b*) ROC curves from '2 K' compressed hand series. Both series of cases were viewed at the start of the study, and again at its end. ■, 1st viewing. ▨, 2nd viewing.

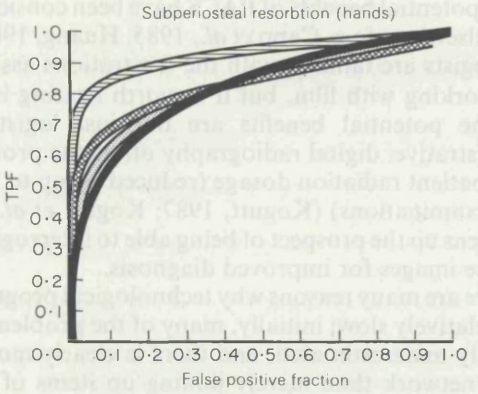


Fig. 3 – ROC curves for pooled observations. Image format: ▣, film; ▨, '1 K' uncompressed; □, '2 K' uncompressed; ■, compressed.

Table 2 – True positive fractions at false negative rates of 15%. Subperiosteal resorbtion: correct diagnosis (raw data)

	Normal ($n = 160$)	*Abnormal* ($n = 160$)	*Total* ($n = 320$)	TPF at FPF = 15%
Film	158 (98·7%)	108 (67·5%)	266 (83·1%)	86·4%
2 K	158 (98·7%)	89 (55·6%)	247 (77·2%)	72·1%
2 K Compressed	155 (96·9%)	70 (43·7%)	225 (70·3%)	63·2%
1 K	156 (97·5%)	73 (45·6%)	229 (71·6%)	67·6%

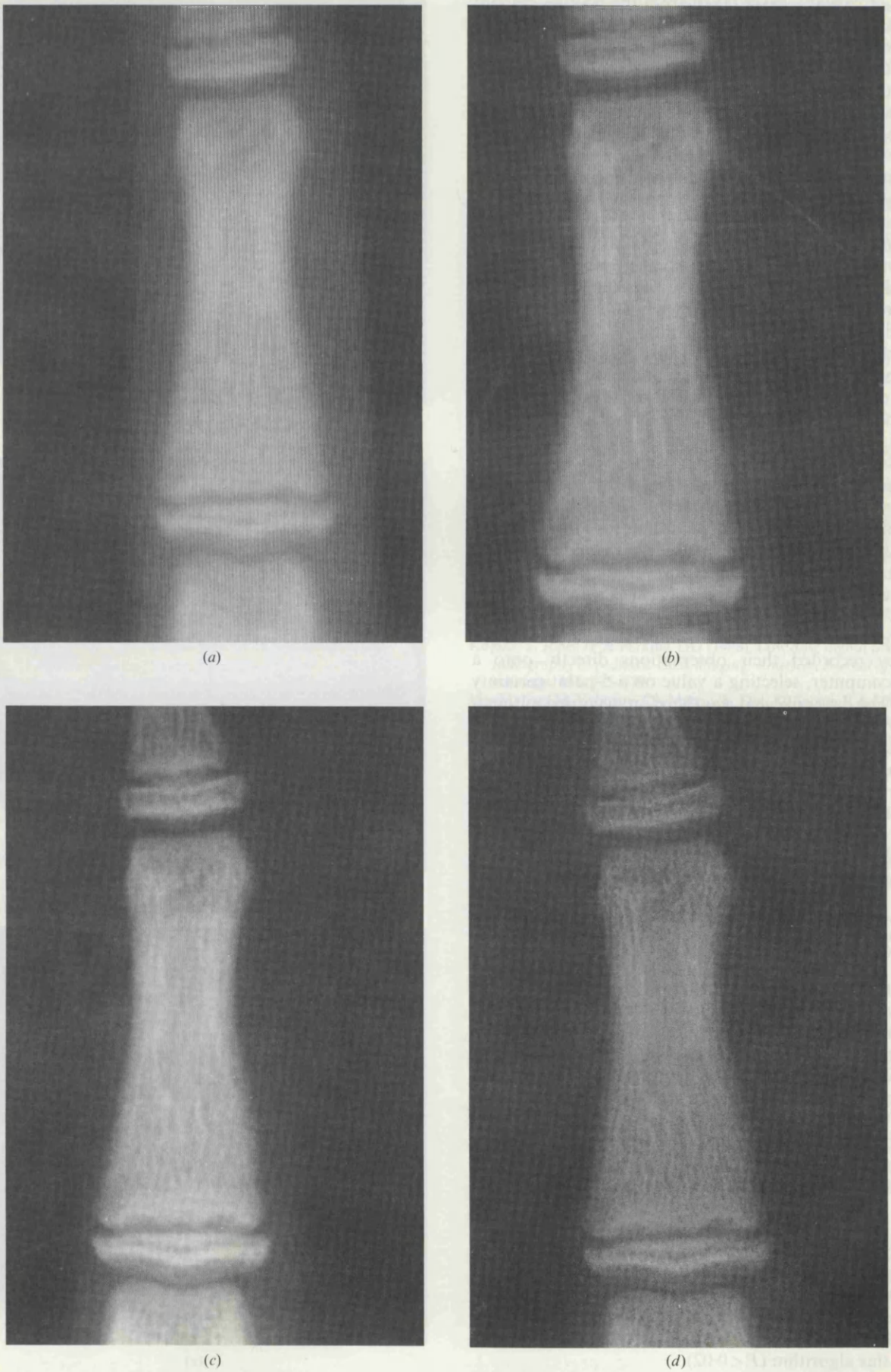


Fig. 4 – Case 1. Patient with long-standing chronic renal failure. (*a*) '1 K'. (*b*) '2 K' with data compression. (*c*) '2 K', no compression. (*d*) Film. Note the raster lines, and the 'blocky' artefact introduced by compression: 5/8 observers made the correct diagnosis on film; 4/8 at '2 K'; and 3/8 on the others.

series of casualty cases taken from the routine casualty workload, 20 in all of which 10 were abnormal, with a diagnosis clearly evident on film. They were shown at '2 K' with compression, as training material at the start of the study, and once more at the end of the main evaluation, to estimate any change in performance over the period of the study.

For the main evaluation study, each viewing session consisted of all 40 images in one of the formats, and the main part of the study therefore consisted of four viewing sessions for each of the observers.

In addition, we wanted to find out if the observers were consistent in their responses; so the first series of displayed hand images that the observers had seen (the '2 K' compressed images) was shown again to all of the participants at the end of the study.

Each observer was required to arbitrate only on the presence or absence of the feature under study. In the case of the casualty images, they were asked to arbitrate on the presence or absence of a fracture. In the case of the renal hands, they were asked to arbitrate on the presence or absence of subperiosteal resorbtion.

Observers were able to manipulate window width and window level using simple control keys, and to select magnification by factors of two or four (two times only, in the case of the 'I K' images). Observers could also rotate the images in 90 degree steps; however, the effect of image orientation in relation to the raster lines was not studied. Apart from data compression, no other image processing features were available on this apparatus.

They recorded their observations directly onto a microcomputer, selecting a value on a 5-point certainty scale. The data collected was therefore recorded automatically in a form that could later be used for statistical analysis.

The results of the observations were evaluated by constructing ROC curves of the pooled data, and the significance of the differences between them was assessed using paired analysis of the parameters of the curves (Swets & Pickett, 1986; Metz, 1978, 1986, 1988), with software from the University of Chicago (Metz, 1987).

RESULTS

Figure 2 shows the ROC curves for the two image series that were shown twice—the casualty cases, and the '2 K' compressed images of the hands.

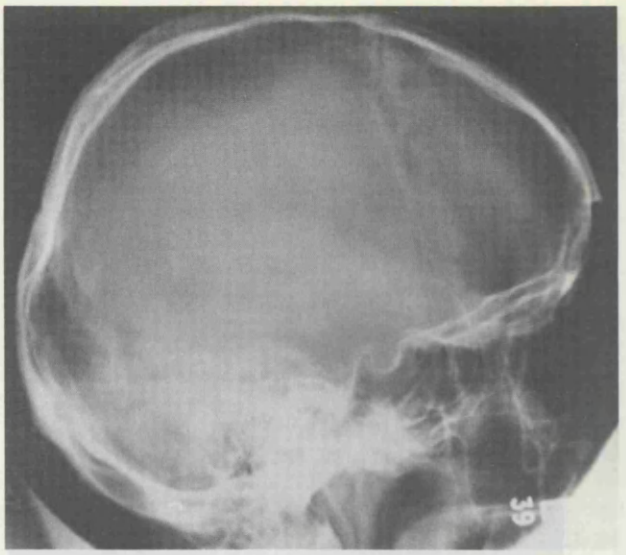
In the case of the fractures, the curves are virtually identical. In the case of the hands, the slight apparent difference between the curves is not statistically significant. These results show a high degree of consistency and reproducibility between the first and second viewing sessions for both series.

Figure 3 shows the ROC curves for the main study. Table 2 shows the true positive fraction at a false positive rate of 15%, which provides a simple aid to comparison.

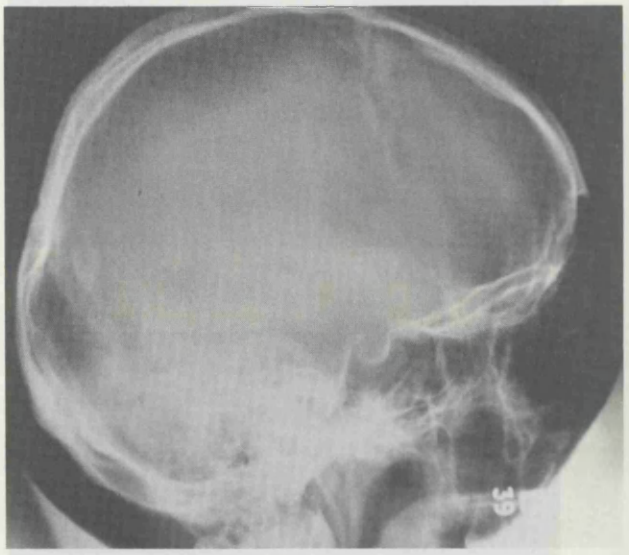
There is a significant difference between the curve for film and the curves for all the VDU displayed formats ($P < 0{\cdot}005$).

There is significant difference between the curve for '2 K' images, and the curve for '2 K' compressed images using this algorithm ($P < 0{\cdot}02$).

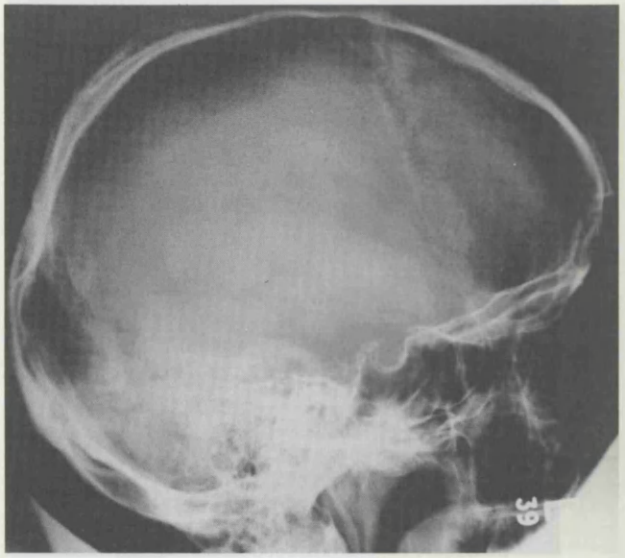
We were unable to demonstrate a significant difference between the '1 K' and '2 K' images, or between the '1 K' and '2 K' compressed images.



(*a*)



(*b*)



(*c*)

Fig. 5 – Case 2. Skull fracture. (*a*) '2 K' with data compression. (*b*) '2 K', no compression. (*c*) Film. This extensive parieto-temporal fracture was missed on four occasions, when observed at '2 K' with compression.

Examples of some of the images are shown in Figs 4 and 5. It is important to note, particularly with the pathological condition selected for study, that the loss of quality in photography and reproduction may well be of the same order as the image degradation in digitisation and display that our study was attempting to assess.

DISCUSSION

We attempted to evaluate the diagnostic image quality of digitised film displayed on a commercially available VDU. We found a significant loss of quality between film and the VDU images. We stress that these findings do not necessarily extrapolate to other equipment, and a different configuration is currently undergoing evaluation.

We also found a significant degradation associated with the compression algorithm that we tested. There are many compression algorithms in existence, and we consider that they each require careful evaluation.

In preparing and conducting this study, we have developed and refined methods that yielded consistent results and will facilitate our evaluation of future installations.

We believe that subjective impressions of displayed images can be misleading, and careful evaluation of diagnostic image quality, using ROC techniques and involving diagnostic radiologists, is an indispensable part of the process of implementing the PACS ideal that must precede any large scale deployment of such systems.

Like St Mary's, departments planning for the future are faced with a dilemma. There are penalties associated with being at the forefront of any field, not least one that involves expensive, high technology equipment. At present, as far as we are aware, no manufacturer has yet installed a complete PACS network. There are problems in interfacing equipment from different manufacturers and even within the range produced by a single manufacturer. New systems need careful evaluation, not just for image quality.

We believe that implementation of the PACS concept is desirable, attainable, and inevitable, and much progress has already been made (Templeton *et al.*, 1987; Taira *et al.*, 1988). It is important for departments under construction today to consider making an investment in this technology even though PACS is not yet fully developed and has not yet been fully evaluated. Clinical evaluation of individual system performance is however of overriding importance prior to commitment to a total system.

Acknowledgements. Our thanks are due to the DHSS, who have founded and supported this project from its inception. We also thank Philips Medical Systems, for their help, cooperation and positive attitude which has contributed to the success of the initial phase of this project. New equipment of higher specification has recently been installed at St Mary's Hospital, and the next phase of the evaluation project is currently under way. We also thank Dawn Beaumont, of the computer unit at St Mary's Hospital, for preparing and processing our data; and the Department of Medical Photography at St Mary's Hospital for preparing the illustrations. We are also grateful to the many people who permitted us to use case material for this project, and particularly to Dr Hugh Saxton and Mrs Mary Robinson of the Department of Radiology at Guy's Hospital. RMD is supported by a grant from Philips Medical Systems, and AWP is supported by a grant from the DHSS. This paper was presented at the Annual Scientific Meeting of the Royal College of Radiologists, Exeter, September 1988.

REFERENCES

Capp, MP, Roehrig, H, Seeley, GW, Fisher, HD & Ovitt, TW (1985). The digital radiology department of the future. *Radiologic Clinics of North America*, **23** (2), 349–354.

Carterette EC, Fiske RA & Huang, HK (1986). Receiver operating characteristic (ROC) evaluation of a digital viewing station for radiologists. *SPIE Proceedings*, **626,** 441–446.

Chakraborty, DP, Breatnach, ES, Yester, MV, Soto, B, Barnes, GT & Fraser, RG (1986). Digital and conventional chest imaging: a modified ROC study of observer performance using simulated nodules. *Radiology*, **158,** 35–39.

Cox, GG, Templeton, AW & Dwyer, SJ (1986). Digital image management: networking, display and archiving. *Radiologic Clinics of North America*, **26** (1), 37–54.

Craig, JOMC (1985). Diagnostic radiology without films. *The Practitioner*, **229,** 1011–15.

Craig, JOMC (1988). The filmless department. *British Journal of Hospital Medicine*, **40,** 97–101.

Huang, HK (1987). *Elements of Digital Radiology: a Professional Handbook and Guide*. Prentice Hall, New Jersey.

Kangarloo, H, Boechat, MI, Barbaric, Z, Taira, RK, Cho, PS, Mankovich NJ *et al.* (1988) Two Year clinical experience with a computed radiography system. *American Journal of Roentgenology*, **151,** 606–608.

Kogutt, MS (1987). Computed radiographic imaging: use in low-dose leg length radiography. *American Journal of Roentgenology*, **148,** 1205–1206.

Kogutt, S, Jones JP & Perkins, DD (1988). Low dose digital computed radiography in pediatric chest imaging. *American Journal of Roentgenology*, **151,** 775–779.

MacMahon, H, Vyborny, CJ, Metz, CE, Doi, K, Sabeti, V & Solomon, SL (1986). Digital radiography of subtle pulmonary abnormalities: an ROC study of the effect of pixel size on observer performance. *Radiology*, **158,** 21–26.

Metz, CE (1978). Basic principles of ROC analysis. *Seminars in Nuclear Medicine*, **8,** 283–298.

Metz, CE (1986). ROC methodology in radiologic imaging. *Investigative Radiology*, **21,** 720–731.

Metz, CE (1987). ROCFIT software package, University of Chicago.

Metz, CE (1989). Some practical issues of experimental design and data analysis in radiological ROC studies. *Investigative Radiology*, **24,** 234–235.

Sakuma, H, Takeda, K, Hirano, T, Okuda, Y, Matsumura, K, Yamaguchi, N *et al.* (1988). Plain chest radiograph with computed radiography: improved sensitivity for the detection of coronary artery calcification. *American Journal of Roentgenology*, **151,** 27–30.

Seeley, GW, Fisher, HD, Stempski, MO, Borgstrom, M, Bjelland, J & Capp, MP (1987). Total digital radiology department: spatial resolution requirements. *American Journal of Roentgenology*, **148,** 421–426.

Spackman, TJ & Bensman, KW (1987). Development of a useful picture archiving and communications system. *American Journal of Roentgenology*, **148,** 1025–1026.

Swets, JA & Pickett, RM (1986). *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory*. Academic Press, London.

Taira, RK, Mankovich, NJ, Boechat, MI, Kangarloo, H & Huang, HK (1988). Design and implementation of a picture archiving and communication system for pediatric radiology. *American Journal of Roentgenology*, **150,** 1117–1121.

Templeton, AW, Cox, GG & Dwyer, SJ (1988). Digital image management networks: current status. *Radiology*, **169,** 193–199.

Templeton, AW, Dwyer, SJ, Cox, GG, Lee, KR, Johnson, JA, Martin, NL *et al.* (1987). A digital radiology imaging system: description and clinical evaluation. *American Journal of Roentgenology*, **149,** 847–851.

Radiology about to go digital

Exciting new technology must be carefully evaluated

Computed tomography, radionuclide scanning, digital subtraction angiography, and magnetic resonance imaging account for up to one third of examinations in modern radiology departments: they yield digital data, giving measurements of physical values from which images are constructed. At present these images are printed on film. The technology now exists to transform other examinations into digital procedures. Computed radiography is a Japanese innovation that replaces *x* ray film with a reusable phosphor material.[1,2] After exposure with conventional radiographic equipment the phosphor sheet is removed from its cassette and scanned to convert the stored image into digital data. The system is more sensitive than film, and more tolerant of incorrect exposure; the procedure entails less radiation and, theoretically, no chemicals and no film.

Linking imaging technology in a computerised network that would allow radiologists and doctors in clinics and operating theatres to deal exclusively with digital images on television systems has long been a pipe dream.[3,4] In the early days, however, few people fully appreciated the true scale of the undertaking. The London Stock Exchange now generates 350 megabytes of data from each day's trading; a fully digital radiology department in a typical hospital might easily produce 2 gigabytes—nearly six times more.

Single full sized optical disks are inadequate for storing such huge amounts of image data. Optical jukeboxes have been devised to handle 90 or more disks, and there is now the prospect of optical tape that might store a terabyte of data (1 million megabytes) on a single reel. In a hospital with a fully digital radiology system thousands of images from hundreds of patients would need to be instantly available for viewing in dozens of locations. This would give rise to enormous problems with data traffic. Laying down standards and finding ways for existing and future equipment—even from a single manufacturer—to communicate in the same language is difficult and costly. The many technical problems have been studied intensively, mostly in the United States.[5,6] The work has been largely conducted by psychiatrists, computer scientists, and engineers.

As solutions appear for technical problems formidable pressure to implement the new technology becomes inevitable—rather like the pressure to introduce a wonder drug that is not yet available for clinical use. This is of special concern to hospitals that are rebuilding or planning for the future and have to consider the prospects for installing a complete network. They need to exercise caution and restraint. There is a sharp contrast between the missionary zeal of those who are frustrated with film but have had no practical experience with the new systems and the realism of established research teams.

system anywhere. There is a strong case to be made against a department or hospital trying to "go digital" in a single move. Even today the text based computerisation of a hospital department, an operation of trivial proportions by comparison, may precipitate chaos. Few hospitals have yet done on a larger scale with text what we would like to do with images. Furthermore, the new equipment is undergoing rapid development: most hardware is obsolescent in about a year.

The most crucial issue is whether the diagnostic quality of the images will be impaired. The average casualty *x* ray department carries out around 140 examinations of normal skulls for each radiograph that shows a fracture; there is no point in being able to store electronically, manipulate, and retrieve all those normal images unless we can be confident that they are truly normal and that the solitary abnormal case will not be missed. Clinical evaluation is now essential—it must include input by diagnostic radiologists and be independent and scientifically sound. Image quality should not be assessed subjectively; exacting scientific methods now permit accurate and reproducible study.[7-9] It is greatly to the credit of the Department of Health that a British department is among the first to have produced an objective appraisal of

Such studies show that much refinement is still necessary.

Unless evaluation retains the priority it deserves there is a danger that implementation will be driven by the technology and by commercial considerations. An incremental, phased implementation of the new digital technology based on careful evaluation of each component is essential.

RICHARD M DAWOOD

Honorary Senior Research Fellow,
St Mary's Hospital Medical School,
London W2 1NY

1 Tateno Y, Iinuma T, Takamo M. *Computed radiography*. Tokyo: Springer Verlag, 1987.
2 Kangarloo H, Boechat MI, Barbaric Z, *et al.* Two-year clinical experience with a computed radiography system. *AJR* 1988;151:605-8.
3 Craig JOMC. Diagnostic radiology without films. *Practitioner* 1985;299:1011-5.
4 Capp MP, Roehrig H, Seely GW, Fisher HD, Ovitt TW. The digital radiology department of the future. *Radiol Clin North Am* 1985;23:349-55.
5 Templeton AW, Cox GG, Dwyer SJ. Digital image measurement networks: current status. *Radiology* 1988;169:193-9.
6 Huang HK. *Elements of digital radiology: a professional handbook and guide*. Englewood Cliffs, New Jersey: Prentice Hall, 1987.
7 Metz CE. Basic principles of ROC analysis. *Semin Nucl Med* 1978;8:283-98.
8 Metz CE. ROC methodology in radiologic imaging. *Invest Radiol* 1986;21:720-31.
9 Metz CE. Some practical issues of experimental design and data analysis in radiological ROC studies. *Invest Radiol* (in press).
10 Dawood RM, Craig JOMC, Highman JH, *et al.* Clinical diagnosis from digital displays: preliminary findings of the St Mary's evaluation project. *Clin Radiol* (in press).
11 Dawood RM, Craig JOMC, Highman JH, *et al.* Evaluation of VDU displays in radiological diagnosis: a preliminary comparison of two systems. *SPIE Proceedings* 1989;1091:368-72.

Correction

Radiology about to go digital

A printer's error occurred in this editorial by Dr Richard M Dawood (5 August, p 340). The work has been largely conducted by physicists, computer scientists, and engineers and not by psychiatrists as published.

Printed in Great Britain by Bourne Offset Limited, Iver, Bucks.

Clinical Radiology (1990) **42**, 6–11

Editorial

Digital Radiology – A Realistic Prospect?

Digital radiology has had a long and uncomfortable gestation. Its ultimate goal can be summarized as the creation within the modern radiology department – and indeed within and between entire hospitals – of an efficient, effective, electronic network capable of handling all diagnostic radiological images, obviating the need for conventional film-based radiology.

This article will examine the present status of the technology that is necessary to achieve this, and will consider current issues that relate to implementation.

IMAGE ACQUISITION

Conventional projection radiography accounts for roughly two thirds of the workload in most modern radiology departments, and handling the resulting images represents the single most challenging technical issue in digital radiology.

In the earliest conceptual stages of its development, attention focused upon a variety of devices for direct capture of digital radiographs, usually incorporating an X-ray source, some kind of scanning procedure, and an array of electronic detectors (Fraser *et al.*, 1983). They were expensive, bulky, unwieldy, and manifestly unsuitable for use on sick people. Also they did not take account of the need to make the most of the huge investment in existing conventional radiographic equipment. It is worth bearing in mind that the current NHS replacement cycle for conventional radiographic equipment can be as long as 30 years. Furthermore, the images resulting from these early attempts at direct digital acquisition were not very good.

The most crucial technological advance to date in image acquisition has been computed radiography (CR). This replaces conventional X-ray film with a re-usable radiation-sensitive phosphor storage medium. Following exposure using conventional radiographic equipment, the sheet of phosphor is removed from its cassette and undergoes a laser scanning process that stimulates luminescence of the stored analogue image, enabling it to be converted into digital data (Tateno *et al.*, 1987). Initially developed by Kodak, manufactured in Japan by Fuji, and now marketed by several companies, the system is much more sensitive than film and has much wider exposure latitude. Experience shows that images of consistent appearance can be readily obtained under a wide range of exposure conditions, even under difficult circumstances such as in theatre or in the intensive care unit (Kangarloo *et al.*, 1988). Radiation doses can be considerably reduced – by up to 98% in certain situations (e.g. follow-up spinal views for scoliosis) (Kogutt, 1987) though reductions of 30–50% are more typical; and in theory anyway, the need for chemical processing and film could be eliminated.

There are believed to be 20–30 CR installations in the USA, approximately 200 in Japan, and 15 in Europe. In some circumstances, CR is an end in itself. A hospital in the USA recently installed CR solely because about 10% of its ICU films were not being reported. No X-ray report meant no revenue. Now they use CR to print out two copies of every image, and the system generates a worthwhile profit.

It is important to recognize that CR is not in fact an example of direct digital acquisition – the image is acquired as a shadow, an analogue 'picture' that is digitized by the scanning laser. This same consideration also applies to items of equipment such as the Konica Direct Digitizer (KDD) – a closed, dedicated chest radiography unit that incorporates a photostimulable phosphor.

Direct digital acquisition, as far as projection radiography is concerned, is not yet a realistic option, though it is currently under active investigation. One promising approach involves the use of charge coupled devices (CCDs). CCDs are currently used, for example, in video cameras; they behave as arrays of tiny electronic detectors, each able to capture the information of a single 'pixel'; the ideal radiographic CCD would need to be as large as a sheet of X-ray film – not achievable at present.

Film digitization has only a limited function within digital radiology. Sheets of film are scanned in a manner that is analogous to the laser scanning processes used in CR systems. The resulting digital data corresponds to a map of the optical density of the film rather than a map of stimulated luminescence. Film digitization is used principally in the research arena, as a gateway for images that will be used in evaluation studies. The effort of digitizing an existing film archive would only be justifiable in the context of a fully-functioning, more extensive digital network.

Images derived from projection radiography carry a huge amount of information. Spatial resolution and the amount of data that result from digitization depend upon the digital matrix size; with scanning laser digitization, this is limited by the size of the scanning laser spot. The amount of data is also related to the number of grey levels represented for each point on the image matrix – the bit depth. The dynamic range of radiographic images is very large: more bits per pixel means better contrast resolution, but also more digital data that will need to be processed, displayed and stored.

In the modern X-ray department, many image types begin life as digital data – notably CT, MRI, and radionuclide scanning. Conversion of the images and signals produced by digital fluorography, DSA and ultrasound scanning, from analogue to digital form, is a routine electronic process that poses no particular problem. All of these image types have matrix sizes that are considerably smaller than those necessary for projection radiography. Such images are in fact relatively easy to acquire, handle, display, archive and communicate in digital form.

DISPLAY

It has been wryly observed that the history of modern radiology is in fact a story of declining image quality: the

best radiographs are produced on non-screen fine grain film; but accepted practice is to use rare earth intensifying screens and faster films with low exposures to limit radiation dosage; and to use a broader focal spot in certain circumstances to preserve X-ray tube life or to reduce equipment costs. Radiologists take pride in their carefully nurtured diagnostic skills – observation, interpretation, and deduction; their ability to home in on the unexpected lesion or the subtle incidental finding sets them apart from their non-radiological colleagues. Radiological diagnosis depends upon image display. How much scope is there for further compromise on image quality?

Images on film are usually referred to as 'hard copy'. Images on TV display systems are, to use the same jargon, 'soft copy'. Digital radiology in its fullest concept only makes sense if TV images can substantially replace the use of hard copy throughout the hospital, with 'soft copy' reporting and viewing by radiologists and referring clinicians.

One index of the quality of images that can be displayed on TV systems is the number of raster lines on the TV monitor. Existing monitors used to display CT, MR, and DSA images are perfectly adequate, and usually display up to 625-lines. It is widely accepted that higher specification is necessary for satisfactory display of radiographic images. Higher specification, however, does not come cheaply.

Designing a suitable high resolution monitor is a costly trade-off between competing technical factors. More lines make it more difficult to achieve refresh rates that avoid flicker in the perceived image. Some monitors use a fast refresh rate in conjunction with an interlaced raster pattern – half of the lines are refreshed at each pass; the interlacing results in a movement effect that can be disturbing to the eye, and that many experts consider unacceptable. Other monitors seek to solve the problem of flicker by using a more persistent phosphor – sometimes causing a visible afterglow between images. There is a 10 to 20-fold difference in image brightness between film on a fluorescent light box and radiographic images displayed on a monitor. This results in reduced grey scale resolution for the displayed images and a critical need for dark viewing conditions; simply increasing the monitor brightness causes blooming of the phosphor and a degraded image – attempts to compensate for this have included the use of fibre optics in the monitor face-plate. The cathode ray tube is not an ideal instrument for the display of stationary, high resolution images, and research efforts are exploring alternatives: among the most promising appears to be the active-matrix liquid crystal system (Kazan, 1989).

Most of the monitors currently proposed for PACS have 1024 or 1280 lines – neither of the two major system vendors currently offers a higher specification. Monitors with 2000 lines have progressed from a research stage to a point where a number of specialist companies are now able to supply them; many of the manufacture and design problems have been solved, but one current sticking point is their satisfactory integration into networks capable of handling the huge amounts of data associated with each image.

A major PACS meeting (Medical Imaging IV, SPIE, February 1990) – at which only a solitary British radiologist was present despite much professed interest in the UK – held a special seminar on high resolution displays. A broad consensus was reached that 1024 and 1280 displays are *inappropriate* for the task of primary radiographic diagnosis in the majority of clinical settings, and that even 2000-line monitors may still not be adequate in their present form. This conclusion accords with work in the UK at St Mary's Hospital, where 1024 and 1280-line monitors have undergone formal evaluation (Dawood *et al.*, 1989a, b).

Monitors of lesser specification than 1024 or 1280 lines have also been proposed, principally to permit review of images outside the main X-ray department ('secondary diagnosis'). At the same meeting, this approach was called into question (H. Kundel, University of Pennsylvania): can we really supply such images to our clinical colleagues on the understanding that they should *not* be used as a basis for important clinical decisions?

Objective assessment of image quality on high resolution displays is an issue of the utmost importance. Many of the physical measurements and methods that can be applied to an image recorded on a sheet of film have no direct counterpart. And there are many psychophysical differences between the task of interpreting an image on film and a displayed image (Kundel, 1986; Kosslyn, 1989). Receiver Operating Characteristic (ROC) methods have been developed and refined to address some of the resulting problems (Metz, 1978, 1986, 1988; Swets and Pickett, 1986; Chakraborty, 1989); they focus upon the diagnostic performance of the displayed image rather than any single physical parameter. ROC studies provide an objective means of comparing performance of different display systems with each other and against film. They have their drawbacks, however: they are difficult and time-consuming to run, and typical ROC tasks—such as determining the presence or absence of a pre-defined lesion from a series of similar images of proven pedigree—do not adequately simulate the real-life radiological task of making a difficult diagnosis from scratch. Subjective assessment of display quality – 'eight out of 10 radiologists found these images acceptable' – is a similar approach to that used in TV petfood commercials, and is not considered appropriate.

Abandoning hard copy reporting represents a fundamental change of practice that can only be justified after the most searching exploration of the consequences for radiological diagnosis. Such work is still in its infancy, partly because the technology has only recently reached a state of sufficient maturity to attract the active participation of diagnostic radiologists; objective assessment is an irksome process that cannot always be in harmony with commercial pressures and consumer aspirations, but we cannot do without it.

Worldwide, almost every institution that has so far installed CR equipment still chooses to print out each image as hard copy. The CR hardware includes a laser printer for sheet film, and these hard copy digital images are at present used for primary diagnosis. This is not quite as wasteful as it seems: square foot for square foot, suppliers have so far held the cost of laser printer film to roughly that of X-ray film; the laser printer images are minified, however, so that the film costs of a CR hard copy service are indeed less. The fullest implementation of digital radiology would allow printing of digital images onto film only in special circumstances, such as for patients being referred to 'non-digital' institutions; however, the 'filmless hospital' may ultimately prove to be as much of an illusion as the 'paperless office'.

IMAGE PROCESSING

Once in digital form, data processing techniques can be applied to enhance, manipulate and interrogate the resulting images. Such processing should be distinguished from simple manipulation of window level and width. Examples include techniques such as subtraction, edge enhancement, noise reduction, contrast equalization, and dual energy filtration, but experience suggests that enhancement of one feature is often at the expense of another (Oestmann *et al.*, 1989). Existing TV display systems inevitably introduce image degradation: it remains to be demonstrated whether image processing methods and window manipulation can adequately compensate for the loss of quality, and can indeed contribute to improved diagnosis.

DATA STORAGE

As a data storage medium, film is costly, bulky, and requires manual filing and retrieval – a tedious process that is notoriously vulnerable to error. What is more, a sheet of film can only be in a single location at any one time. In a single exposure, the amount of information that can be captured is restricted by the latitude of the film. Such limitations have provided powerful impetus to the development of digital systems.

Image data storage requirements for an all-digital 800-bed hospital have been estimated to be at least 2 Gbytes (2000 Megabytes) per day. Magnetic hard disks allow rapid access times, but have a limited capacity; they are costly, but data can be continually over-written; their most useful role will ultimately be the provision of rapid-access buffer storage rather than a permanent archive. In December 1989, IBM announced a new method of data storage, increasing the density of magnetic hard disk storage by a factor of 15–30 with no loss of speed (personal communication, Dr K. Keeshan, IBM, San Jose, USA). This may have important future implications for PACS.

Optical disks store digital data at higher density. They are relatively cheaper, and a little slower. Storage is permanent, and the data cannot readily be over-written. Fourteen-inch conventional disks can store 6 GBytes per disk, though this is still not enough to enable single disks to provide adequate on-line storage: optical jukeboxes have therefore been devised to handle 90 or more disks at a time. Widespread application of compact disk technology in the domestic audio field has so far provided the greatest stimulus to development and refinement of optical disk technology, and the manufacturing costs of compact disk drives are now small. It remains to be seen whether there will be any benefits in terms of improved design and reduced cost of optical jukeboxes, which are at present cumbersome, mechanical and unwieldy. Current models can cost as much as £500 000 to install; 'mirroring' data, for extra security, may necessitate a second, back-up archive – and some would deem this an essential requirement.

There has been a recent new development that appears to hold much promise. It is called optical tape, and has been developed in the UK by ICI. It uses optical storage technology with higher capacity at a fraction of the cost. A single 12-inch reel of optical tape is capable of storing 1 Terabyte of data – a million Megabytes, probably equivalent to more than 1 year's image output from an 800-bed hospital. Access times are relatively slow – 58 seconds from one end of a TByte tape to the other (and proportionately less with shorter tapes). The film cost of a chest X-ray is about £1.50 per sheet; storing the same image on optical tape costs about 1.5 pence. The first optical tape drives (CREO Systems) are currently under evaluation in Canada, and are planned for supply in Europe within the next 12 months (personal communication, D. Bennett, ICI Imagedata, UK). Drive cost is likely to be around £120 000, but a viable archive would need more than one. The prospect of being able to set up a number of duplicated image archives at various points on a PACS network is alluring, but it remains to be seen whether this medium will truly meet the high expectations that are currently held for it.

Other storage media are also worthy of mention. Optical tape technology is also being applied to the development of optical floppy disks (Bernoulli Systems), capable of storing 1 GByte at low cost. A piece of tape the size of a credit card would be capable of storing between 150 and 300 MBytes; at a cost of pennies, it could reasonably be issued to patients as a portable, back-up copy of their own images. A prototype system based on 3.5 inch WORM compact optical discs (capacity 180 MBytes, and more expensive) has already been successfully introduced on a small scale at UCLA to do just this (Cho *et al.*, 1990). Re-writable magneto-optical disks have been developed, and prices will soon fall dramatically. Such developments will make it possible to download copies of image data for clinics or selected departments, relieving demand on the main network, or perhaps reducing the need to extend the network far from the radiology department.

Storage capacity can be increased by means of data compression techniques. It is possible to encode data with no loss of information at compression ratios of up to about 3:1. Further compression may result in slight data loss, but it seems likely that clinically useful compression ratios of up to 20:1 may be feasible. These so-called 'lossey' algorithms require careful evaluation for diagnostic image quality using ROC techniques. A number of manufacturers, notably in the USA, have expressed deep concern about the medico-legal implications of wilfully applying techniques known to impair image quality, and have focused their anxieties upon data compression in particular, though such considerations might equally apply to many other aspects of this evolving technology that result in images of lesser resolution than is currently achievable with film.

NETWORK FUNCTIONS

Connecting up individual system components is the easiest task in the world, but only on paper: all it takes is a few pen strokes on a diagrammatic plan. In reality, computerized networks are complex, fragile systems that demand the utmost care in setting up and supporting. The network problems have absorbed much of the research effort that has so far been invested in PACS, both by manufacturers and by individual institutions (Templeton *et al.*, 1988).

Linking and interfacing different types of equipment, and finding ways for existing and future equipment to communicate in the same language is a fearsome task. Standards such as ACR/NEMA (American College of

Radiology/National Electrical Manufacturers Association) and extensions to it such as SPI (Standard Product Interchange) are designed to overcome some of the connectivity problems, and are at an advanced stage of development (Horii *et al.*, 1990). They are costly and can reduce performance of existing equipment, and they do not yet provide a complete solution.

In a fully operational digital hospital, the network would have to be able to satisfy requests to distribute images belonging to any patient, to any location in the hospital: this would mean that thousands of images from hundreds of patients would need to be instantly accessible for viewing. The data traffic problems and the practical problems of data management, are enormous, though they are not necessarily the same in every country: in Sweden, for example, accepted procedure is that images do not leave the X-ray department; a Swedish network would presumably be easier to run than one at St Mary's Hospital, London, where it was at one time considered that a total system would require 140 separate image display workstations. A further problem is speed: the amounts of data involved are potentially huge, and the systems so far available are disappointingly slow. Data compression speeds data transmission, and is one partial solution; fibre optic cable links on a network can further increase speed – though they add further to system costs. Experience with a recent clinical installation in Italy designed to handle CT images, shows that reporting times have increased by more than 30%, to a point where radiologists feel a powerful disincentive to use it (Ukovich *et al.*, 1990). Experience with the research installation at St Mary's shows a three to fourfold increase in reporting times for displayed images compared with film radiographs, with well-trained observers performing identical tasks. Ultimately, no radiologist will agree to work with a system that intrudes significantly upon his or her productivity, regardless of how useful the images might be.

Another obstacle is the problem of linking text to the images through the radiology information system (RIS), and providing adequate communication with hospital information system (HIS) and patient databases. For example, the HIS will have an advance record of all patients attending a particular clinic; the RIS will have a record of those patients who have had radiological investigations; collating and transmitting the relevant images and reports – perhaps to the clinic's own database – should be an automatic and effortless process performed at 'off-peak' times.

A fully functioning system offers the prospect of image communication and transmission not just across a hospital network, but between different centres – teleradiology. This is of particular benefit where distances are great and specialist skills are scarce. At a number of centres in the USA, for example, neuroradiologists are already able to monitor patients undergoing CT scanning at clinics many miles away, and some centres even have their own satellite links. Transmission over the public telephone system is relatively slow and susceptible to degradation; while it is unsuitable for a large volume of image data, CT and MR images can be transmitted quite adequately by this route. Perhaps of more immediate practical importance in the UK, radiological reports can already be transmitted to referring GPs own computers over the phone system.

Data security is an important pre-occupation in business and industry. In networks such as airline reservation systems, it has taken years to refine. Data must be protected from unauthorized access and misuse as well as from inadvertent corruption, and in PACS too this will inevitably demand careful consideration.

WORK PRACTICES

What will be the effect of digital radiology upon the way we work? How will the relationship between radiologists and clinicians change? Should each image become accessible on the network as soon as it is taken, or only after it has been reported? Image transmission for case conferences between different institutions – perhaps even in different countries – could become commonplace. And if our clinical colleagues want another opinion, they could conceivably shop around by satellite, perhaps. . .

While it is intriguing to speculate on some of these issues, it may also be of limited value – much will have to depend on eventual system configuration, performance and practical operating constraints.

IMPLEMENTATION: BIG BANG, OR BIG DISTRACTION?

It remains true that no institution in the world has yet achieved the target of a functioning, all-digital radiology department. Looking dispassionately at the present state of play, it is difficult to feel confident about the performance of a total system that would have to be constructed from currently available equipment; an impromptu poll of delegates at a recent international meeting found hardly anybody willing to accept the likelihood of viable total systems before 1995.

Some may find this disappointing. After all, digital radiology has now been under development for almost a decade, and the original concept seemed simple enough; its visible progress has hardly matched the brisk advances that have taken place in other areas of medical imaging over roughly the same period – with ultrasound, CT, gamma cameras and new radio-isotopes, new radiographic contrast media, DSA and MRI. If there is an element of failure implicit in its slow progress, it is above all the initial failure to fully recognize the sheer technical enormity of the task that was being addressed.

Some may also find it difficult to be dispassionate about the need for implementation – there is an overwhelming temptation to believe that digital systems will somehow provide a blanket solution to all of the daily frustrations that go with providing a radiology service in today's overstretched and chronically under-funded hospitals, including constant trouble with lost or missing films, and with film filing, handling and administration.

It *is* clear that there is a powerful and irreversible movement of modern medical imaging technology towards digital systems, but it is a fallacy that implementation in a single step is essential or even remotely desirable. The technology is now hardly in an optimal state and is at present undergoing rapid evolution. This is a necessary, ongoing process, that can only be exploited to full advantage by a rolling programme of implementation that would allow networks to grow over time, with gradual redeployment of obsolescent equipment in an imaginative manner away from the most sensitive points on the network (Dawood, 1989a). While not a view that is

universally shared, this approach is widely held by those with practical experience of PACS. Experimenting with a total system would be an interesting exercise if cost and patient care were no object, and the system could be regarded as utterly disposable.

The pressure to install total digital systems in a single move appears to be a predominantly British phenomenon, for two main reasons. Most radiology departments do not have a programme of capital equipment expenditure able to meet conventional replacement needs, let alone the cost of new digital systems; it is much easier to obtain funding for new equipment by incorporating the costs into the considerably larger construction and commissioning budgets of a new hospital or wing. Secondly, economic justification is perceived by some to be essential, and the most impressive predicted savings only appear when calculations allow total elimination of costs that attend the use of film.

Hospitals currently planning for redevelopment therefore face perplexing choices, apparently between two incompatible extremes. The opportunity to 'go digital' is viewed as one that may not come again, but is attended by a host of unknown quantities – technical viability and clinical effectiveness among them – that have to be reconciled with the need to provide a good clinical service from the first day of operation. Alternatively, failure to grasp this challenge may mean confinement to conventional, film-based systems with no further prospect of introducing digital technology until the day that the newly built department eventually falls down. The absence of any comprehensible policy on PACS from the Department of Health, and the prevailing feeling that some of the hospitals currently examining PACS feasibility are competing for the same funds, does not make the choices easier.

It is not the purpose of this article to seek to justify PACS in terms of cost; all kinds of statistical manipulations have been proposed to do so, and one computer model uses 500 variables. There are too many uncertainties to make any meaningful conclusion possible. The idea that a filmless system is necessarily a cheaper one is fanciful – except perhaps in the very long term. It has long been realized that low-paid film filing clerks will have to be replaced by highly skilled computer staff and medical physicists (Gray *et al.*, 1984). There will always be a need to produce certain images as hard copy – no-one knows how many. Digital storage media – disks or tape – have a significant cost; and of course the digital equipment itself has a high capital cost, over which obsolescence looms almost from the day that the order is placed. Equally, it is unreasonable to judge the benefits purely in economic terms; some would argue that the prospect of a smoothly running and fully effective, integrated electronic X-ray department, with reduced radiation doses for patients and staff, is beyond price.

Nor is it the purpose of this article to argue the case for proceeding with implementation: the merits of PACS have been reviewed quite adequately by other authors (e.g. Capp *et al.*, 1985; Craig, 1985; Huang, 1987).

If PACS is to proceed at all in the UK, however, it will have to be properly funded. The NHS has so far been insulated from many of the true costs of modern imaging developments: CT and MR scanners that in other countries are the normal, accepted responsibility of the prevailing health care system, have had to be obtained by appealing to public charity. PACS cannot be funded by public appeal: it is not one single item of hardware, but an investment in the very fabric of hospital infrastructure, that will require a continuing commitment to support and maintain. It cannot be settled in a single step with a one-time payment. Should the peculiar requirements of this new technology be forced to adapt to the rigid structure of traditional approaches to funding, or should the reverse happen?

Once a decision has been taken to support a PACS installation, funds should be secured for implementation at a rate that reflects current progress and success with each stage. Funding bodies that lack the foresight to consider such flexibility of approach will inevitably otherwise face elaborate and wasteful attempts to mislead them that total systems are indeed the only way that implementation can proceed.

CONCLUDING COMMENTS: PRACTICAL IMPLICATIONS

What can a hospital that aspires to go digital do now? Acquisition techniques in general radiography – such as CR – are likely to remain compatible with existing and future general radiographic equipment, so that no special provision needs to be made here. New CT, MR, digital fluorography and DSA installations should be specified to incorporate (or be readily adaptable to) the latest version of standards such as ACR/NEMA and SPI; image quality issues are largely irrelevant, and a number of institutions have begun to establish prototype networks to handle these images, with promising results. CR is now sufficiently mature to justify controlled implementation for further study and evaluation in areas where the benefit is likely to be greatest, such as for portable work and in the ICU; hard copy output will be necessary, but it is worth installing a display system on a small scale for evaluation.

No less important is preparation of staff, for whom continuing education, dialogue with other centres, and attendance at conferences and meetings are essential. The objective must be to build up an infrastructure of supporting staff – especially radiographers and medical physicists – with the necessary skills and experience to be able to implement larger systems successfully in due course. It is also important to begin correcting operational inefficiencies without delay: PACS will not cure poor motivation, sloppy administrative practices and bad technique – it may make them worse – and it will certainly bring problems of its own that have not yet been thought of.

New buildings should allow for ducting, air conditioning and power requirements – though most new hospitals have these facilities anyway. They should also have a layout able to cope with the co-existing needs of conventional and digital systems as implementation progresses. Choice of equipment from a constantly expanding pool is difficult – manufacturers are persuasive, and anxious to recover their substantial development costs; it is helpful if staff within a department possess the skills necessary to assess proposals objectively. And in discussions with suppliers, bear in mind that it is not necessary to purchase every item of equipment on the network from the same source. Items from the same supplier are not necessarily from the same manufacturer, and a single source is not a guarantee of compatibility. Some suppliers offer deals in

which conventional radiographic equipment and digital items are included together as a package. Do not be diverted from the course of rational implementation by attractive-sounding special-offers, and make sure that the package includes only equipment that you really want and would have chosen anyway, conventional or digital.

In most other countries, total systems are not even an issue. The institutions that have made most progress towards implementation have been established departments that have introduced new technology step by step. They have not been purpose-built. They have ongoing programmes of research and evaluation. They have relied upon no single manufacturer or supplier as exclusive sources of hardware or of know-how, and have devoted considerable effort to developing support skills in-house: at UCLA, for example, the digital unit now has a physics and scientific staff of almost 40 people. They have carefully cultivated links with academic and scientific departments at nearby universities. And they have generally had access to government funding and research grants. Not surprisingly, it takes hard work.

The technology of PACS is challenging and complex, but must not be allowed to obscure its true function: better patient care and radiological diagnosis. It must be implemented in a rational, scientific, responsible manner that will allow these objectives to be truly achieved. The need for careful evaluation of each step prior to implementation has been questioned: with PACS, like with the world's first airplane, one sooner or later has to take to the air (Adam and Allison; 1989; Hemingway, 1989, Dawood, 1989b). The Kitty Hawk won the race to be first and demonstrated convincingly that flight was possible, but she survived in the air for just 3 minutes before crashing beyond repair: we shall have to do much better than that with PACS.

Will radiology really be filmless? The answer is 'yes', but the question is unimportant. At issue is not the removal of film, but the introduction of advanced information technology into radiology. Total systems will be a long time coming, and are a poor excuse for delaying that process.

Acknowledgements: This article is based on work with digital display systems in the Department of Radiology at St Mary's Hospital, and visits and discussions at numerous research sites abroad. I should like to express my gratitude to Dr Oscar Craig and my colleagues at St Mary's, and also to the many colleagues in the UK and abroad who have contributed to ideas put forward in this article.

REFERENCES

Adam, A & Allison, DJ (1989). Radiology about to go digital (letter). *British Medical Journal*, **299,** 69.

Capp, MP, Roehrig, H, Seeley, GW, Fisher, HD & Ovitt, TW (1985). The digital radiology department of the future. *Radiologic Clinics of North America*, **23,** 349–354.

Chakraborty, DP (1989). Maximum likelihood analysis of free-response receiver operating characteristic (FROC) data. *Medical Physics*, **16,** 561–568.

Cho, PS, Osman, R, Ho, BK, Morioka, CA, Marx, P, Huang, HK, Shinagawa, T & Sonobe, T (1990). Personal digital image filing system. *Proceedings of the SPIE*, **1234,** (In press).

Craig, JOMC (1985) Diagnostic radiology without films. *The Practitioner*, **229,** 1011–1015.

Dawood, RM, Craig, JOMC, Highman, JH, Wadsworth, J, Glass, HI, Todd-Pokropek, A *et al.* (1989a). Clinical diagnosis from digital displays: preliminary findings of the St Mary's evaluation project. *Clinical Radiology*, **40,** 369–373.

Dawood, RM, Craig, JOMC, Highman, JH, Wadsworth, J, Glass HI, Todd-Pokropek A *et al.* (1989b). Evaluation of VDU systems in radiological diagnosis: a preliminary comparison of two systems. *Proceedings of the SPIE*, **1091,** 638–671.

Dawood, RM (1989a). Radiology about to go digital: exciting new technology must be carefully evaluated (leading article). *British Medical Journal*, **299,** 340–341.

Dawood, RM (1989b). Radiology about to go digital (letter). *British Medical Journal*, **229,** 854.

Fraser, RG, Breatnach, E & Barnes, GT (1983). Digital Radiography of the chest: clinical experience with a prototype unit. *Radiology*, **148,** 1–5.

Gray, JE, Karsell, PR, Becker, GP & Gehring, DG (1984). Total digital radiology: is it desirable or feasible? *American Journal of Radiology*, **143,** 1345–1350.

Hemingway, AP (1989). Radiology about to go digital (letter). *British Medical Journal*, **299,** 69.

Horii, SC, Hill, DG, Blume, HR, Best, DE, Thompson, BG & Fuscoe, CA (1990). ACR/NEMA standards activity: an update. *SPIE Proceedings*, **1232** (In press).

Huang, HK (1987). *Elements of Digital Radiology: A Professional Handbook and Guide*. Prentice Hall, New Jersey.

Kangarloo, H, Boechat, MI, Barbaric, Z, Taira, RK, Cho, PS, Mankovich, NJ *et al.* (1988) Two year experience with a computed radiography system. *American Journal of Radiology*, **151,** 605–608.

Kazan, B. (1989). New electronic displays and their possibilities for radiology. *Investigative Radiology*, **24,** 817–819.

Kogutt, MS (1987). Use of a computed radiography system for low-dose digital radiography of infants and children. *Radiology*, **165,** 415–416.

Kosslyn, SM (1989). The psychology of visual displays. *Investigative Radiology*, **24,** 417–419.

Kundel, HL (1986). Visual perception and image display terminals. *Radiology Clinics of North America*, **24,** 69–78.

Metz, CE (1986). ROC methodology in radiologic imaging. *Investigative Radiology*, **21,** 720–731.

Metz, CE (1988). Some practical issues of experimental design and data analysis in radiological ROC studies. *Investigative radiology* (in press).

Metz, CE (1978). Basic principles of ROC analysis. *Seminars in Nuclear Medicine*, **8,** 283–298.

Oestmann, JW, Greene, R, Rhea, JT, Rosenthal, H, Koenker, RM, Tillotson, CL *et al.* (1989). Single exposure dual energy digital radiography in detection of pulmonary nodules and calcifications. *Investigative Radiology*, **24,** 517–521.

Oestmann, JW, Rubens, JR, Bourguin, PM, Rhea, JT & Llewellyn, HJ, Greene, R (1989). Impact of post processing on detection of simulated pulmonary nodules with digital radiography. *Investigative Radiology*, **24,** 467–471.

Swets, JA & Pickett, RM (1986). *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory*. Academic Press, London.

Tateno, Y, Iinuma, T & Takano, M (eds) (1987). *Computed Radiography*. Springer Verlag, Berlin.

Templeton, AW, Cox, GG & Dwyer, SJ (1988). Digital image management networks: current status. *Radiology*, **169,** 193–199.

Ukovich, W, Dalla Parma, L, Stacul, F, Zernich, RC & Girivona, P (1990) Practical experience with PACS: benefits and pitfalls. *Proceedings of the SPIE*, **1234,** (In press).

R. DAWOOD
Honorary Senior Research Fellow
St Mary's Hospital Medical School &
Imperial College of Science,
Technology & Medicine, London W2

Evaluation of VDU displays in radiological diagnosis: A preliminary comparison of two systems

R M Dawood[1], J O M C Craig[1], J H Highman[1], J Wadsworth[2], H I Glass[3], A Todd-Pokropek[4], D A Cunningham[1], J M Stevens[1], A Al-Kutoubi[1], R W Kerslake[1], C J Barber[1], M C Crofton[1], A W Porter[1].

[1]Department of Radiology, St Mary's Hospital, London, U.K.
[2]Academic Department of Community Medicine, St Mary's Hospital Medical School, London U.K.
[3]North West Thames Regional Health Authority, London U.K.
[4]Department of Medical Physics, University College London, U.K.

ABSTRACT

As part of the evaluation of the display requirements needed in PACS, an ROC curve study was performed in order to assess clinical performance and compare two different display workstations using the results from the original films as a reference. It was found that in both cases the ROC curves obtained from the displays were significantly worse than obtained from the original film. It is therefore suggested that considerable care needs to be paid to the performance of such displays for them to fulfil the requirements of a clinically viable PACS system.

1. INTRODUCTION

Picture Archive and Communications Systems (PACS) are of considerable interest for the future of radiology.[1-3] Work at St Mary's Hospital, London, has largely focused upon one single, crucial element of the digital chain, the digital display.[4] The main need from the point of view of diagnostic radiology is to evaluate VDU (CRT) displays in terms of their clinical radiological diagnostic performance. The most important question, is still largely unanswered : will we be able to make the correct diagnosis?

It has now been possible to examine two different, commercially available display systems, and assessed them using the same clinical material and the same team of observers. The basic method used to analyze the data, as has been employed in a number of other trials,[5-8] was to establish Receiver Operating Characteristic curves, for which a well established methodology exists.[9-12]

2. MATERIAL

The two systems used for the comparison are summarized in terms of the ability to digitize films, their local computational and storage facility and their display.

The first system (System 1) comprised:

Two laser film digitizers (scanning spot size approximately 100 microns) configured at two matrix sizes - nominally "1K" and "2K",

An optical disc archive, with a single disc drive (1Gbyte per side) and a 160 Mbyte magnetic disc buffer,

An image management system,

A display terminal consisting of a 4-monitor viewing station with 160 Mbyte magnetic disc for local storage,

1024 line monitors were used, with a 48MHz band width and 8-bit video coding, interlaced.

The second system (System 2) comprised:

Two laser film digitizers, scanning spot size 200 microns, 180Mbyte local storage,

An optical disc archive, with a single disc drive (1Gbyte per side) and 4x825 Mbyte magnetic disc storage,

A display terminal consisting of 2 monitors with 180 Mbyte magnetic disc for local storage,

Two 1280 line monitors with a band width corresponding to a 70MHz bandwidth and 12-bit. video coding, non-interlaced.

The clinical material used for this study comprised a series of 40 hand radiographs. Half were from patients who had no known renal disease, and who had attended the X-ray department for a variety of unrelated conditions; the other half were from patients with proven chronic renal failure in whom a diagnosis of subperiosteal resorption had been evident radiologically on at least two occasions and in whom the diagnosis was also visible to the radiologist administering the tests.

3. METHOD

With System 1, the original films were digitized at a nominal "1K" and "2K" matrix size. To be more precise, for a film size of 24x30 cm, these corresponded to matrix sizes of 750x975 and 1500x2600 respectively. The display matrix size was 1024x780 (approx).

With System 2, the original films were digitized at 200 microns, corresponding to a nominal matrix size of "1K", that is, for a 24x30cm film to a 1200x1500 matrix size. The display matrix size was 1024x850 (approx).

The images were presented to a team of seven observers for interpretation. In one viewing session, an observer viewed all 40 images in one format. Each observer was required to arbitrate only on the presence or absence of the feature under study. In the case of the renal hands, they were asked to arbitrate on the presence or absence of subperiosteal resorption.

Their observations were recorded directly onto a microcomputer, by selecting a value on a 5-point (continuous) certainty scale.

The results of the observations were evaluated by constructing ROC curves of the pooled data, and the significance of the differences between them was assessed using paired analysis of the parameters of the curves with software from the University of Chicago.[13]

In addition to these ROC curve studies, physical measurements were made of these (and other) monitors to determine certain parameters of the display, such as uniformity, temporal stability, flicker, jitter etc. This measurements were made using a fast photodiode connected to either an oscilloscope, or a frequency spectrum analyzer. These results have been reported elsewhere.[14]

4. RESULTS

The results are summarized in Figure 1 which shows the ROC curves for a) the original films b) System 1 digitized at "1K" c) System 1 digitized at "2K" c) System 2 digitized at "1K". The ROC curve for the original film is above the ROC curves for all the displays. The ROC curve for the System 1 "1K" display appears to be the worst.

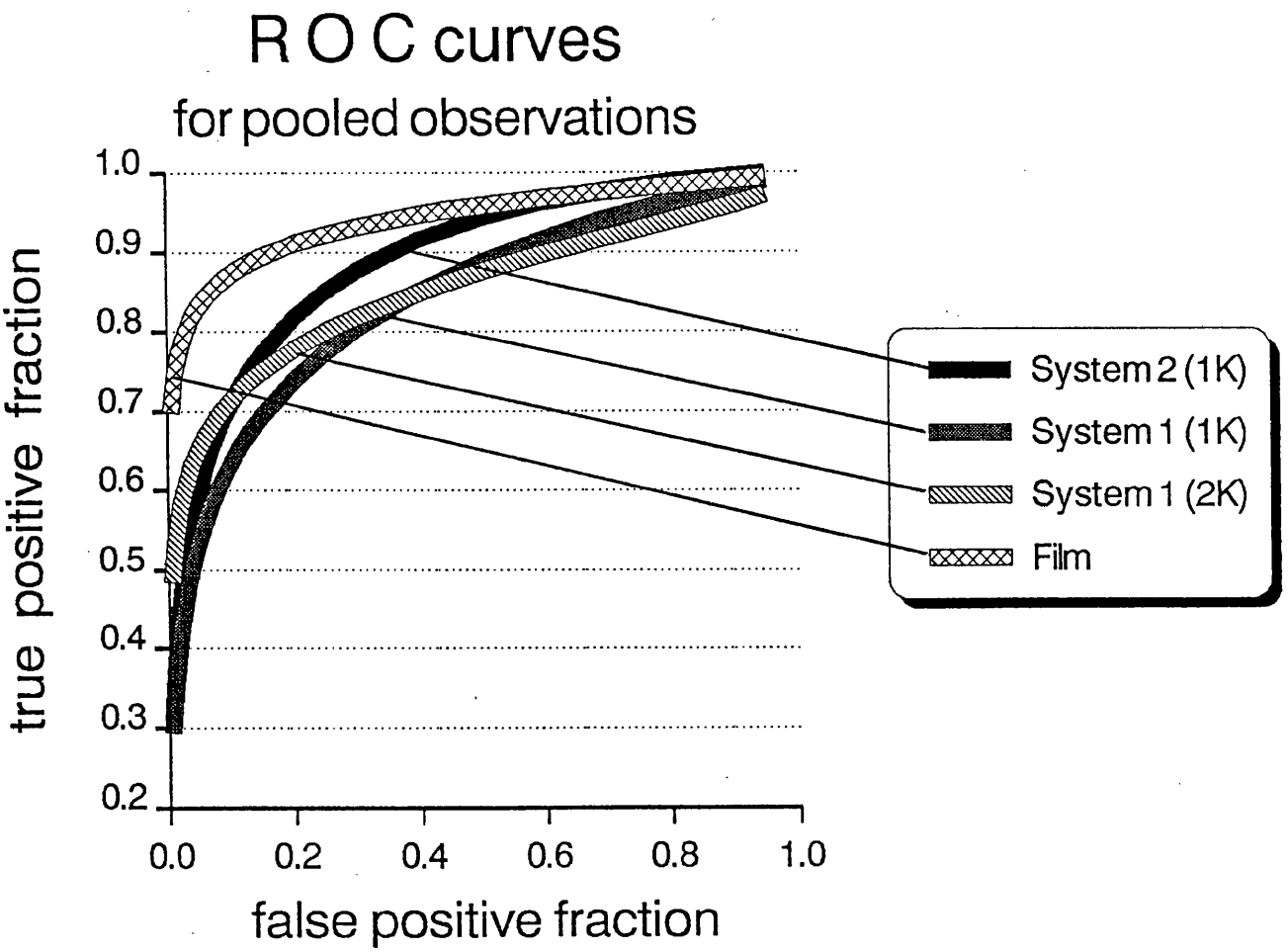


Figure 1
The ROC Curves showing the differences between the interpretation based on the use of the original film and the three types of display.

TABLE 1
TRUE POSITIVE RATE
AT FALSE POSITIVE RATE=15%

Original film	89.4%
System 1: "1K"	69.4%
System 1: "2K"	75.0%
System 2: "1K"	77.3%

To assist the interpretation of these results, Table 1 shows the true positive fraction at a false positive rate of 15%, indicated to provide a simple aid for comparison of these ROC curves. Table 2 shows the significance of the differences between these ROC curves. The interpretation of the film was significantly better than for any of the displays. The display from the System 2 "1K" was better than for either of the System 1 displays although these differences fail to be significant ($p<0.1$).

TABLE 2
SIGNIFICANCE OF DIFFERENCES
OF ROC CURVES

SYSTEM	Original Film	System 1 "1K"	"2K"
1: "1K"	<0.002		
1: "2K"	<0.005	<0.85 [NS]	
2: "1K"	<0.05	<0.1 [NS]	<0.1 [NS]

Expressed in terms of p values
[NS] indicates No Significant difference

5. DISCUSSION

In evaluating the diagnostic image quality of digitized film displayed on two commercially available VDU systems, a significant loss of quality was found between film and the VDU images using both systems.

Subjective impressions of displayed images can be misleading, and careful evaluation of diagnostic image quality, using ROC techniques and involving diagnostic radiologists, is an indispensable part of the process of implementing the PACS ideal that must precede any large scale deployment of such systems. Performing proper physical measurements of the display system are also essential in ensuring adequate image quality.

6. CONCLUSIONS

The major conclusions which this study suggests are that:

1. There are important sources of loss of image quality, in addition to those that result from an inadequate matrix size, which can degrade performance, significant with respect to reporting from film,

2. It is vitally important to perform objective tests, both ROC studies and physical measurements, such as described in this paper.

7. ACKNOWLEDGMENTS

We would like to thank the DHSS, who have funded and supported this project from its inception, Philips Medical Systems, Dawn Beaumont of the computer unit at St Mary's Hospital, and the Department of Medical Photography at St Mary's Hospital.

8. REFERENCES

1. M.P. Capp M P, H. Roehrig, G.W. Seeley, H.D. Fisher and T.W. Ovitt, "The digital radiology department of the future" Radiologic Clinics of North America 23, 2 349-354 (1985).

2. J.O.M.C. Craig, "The filmless department" British journal of hospital medicine 40, 97-101 (1988).

3. J.O.M.C. Craig, " Diagnostic radiology without films" The practitioner, 229, 1011-15 (1985).

4. R.M. Dawood, J.O.M.C. Craig, J.H. Highman, J. Wadsworth, H.I. Glass, A. Todd-Pokropek, D.A. Cunningham, J.M. Stevens, A. Al-Kutoubi, R.W. Kerslake, C.J. Barber, M.C. Crofton and A.W. Porter, "Clinical diagnosis from digital displays: preliminary findings of the St Mary's evaluation project" Clinical radiology, In press (1989).

5. D.P. Chakraborty, E.S. Breatnach, M.V. Yester, B. Soto, G.T. Barnes and R.G. Fraser, "Digital and conventional chest imaging: a modified ROC study of observer performance using simulated nodules" Radiology 158, 35-39 (1986).

6. E.C. Carterette, R.A. Fiske, H.K. Huang, "Receiver operating characteristic (R O C) evaluation of a digital viewing station for radiologists" Proc SPIE 626, 441-446 (1986).

7. H. MacMahon, C.J. Vyborny, C.E. Metz, K. Doi, V. Sabeti and S.L. Solomon, "Digital radiography of subtle pulmonary abnormalities: an ROC study of the effect of pixel size on observer performance" Radiology, 158, 21-26 (1986).

8. G.W. Seeley, H.D. Fisher, M.O. Stempski, M. Borgstrom, J. Bjelland and M.P. Capp, "Total digital radiology department: spatial resolution requirements" American journal of roentgenology. 148, 421-426 (1987).

9. C.E. Metz, "R O C methodology in radiologic imaging" Investigative radiology, 21, 720-731 (1986).

10. C.E. Metz, "Some practical issues of experimental design and data analysis in radiological R O C studies" Investigative radiology, In press.

11. C.E. Metz, "Basic principles of R O C analysis" Seminars in nuclear medicine. 8, 283-298 (1978).

12. J.A. Swets and R.M. Pickett, Evaluation of diagnostic systems: methods from signal detection theory Academic Press, London (1987).

13. C.E. Metz, ROCFIT software package, University of Chicago (1987).

14. R.M. Dawood, A. Todd-Pokropek, J.O.M.C. Craig, J.H. Highman, and H.I. Glass, "Monitor quality and its importance in a PACS display workstation", Radiology 169(P), 358, (1988).

The perception of noise on a PACS display, and its influence on signal to noise ratio on a film digitizer

R M Dawood[1], A Todd-Pokropek[2],
J O M C Craig[1], J H Highman[1], A W Porter[1].

[1]Department of Radiology, St Mary's Hospital, London;
[2]Department of Medical Physics, University College London.

ABSTRACT

The signal to noise ratio and the number of useful bits in data derived from laser film digitization has been assessed for different film densities. The spot size of 210μm was not modifiable, and films were digitized to 12bits of grey scale. It was found that the useful dynamic range was typically less then 256, and that the least significant bits were primarily noise. This study was part of a series of tests that have been performed using resolution patterns, low contrast objects and clinical test series, from which it appears that noise needs to be perceptible in the image for good performance in detecting subtle features. Results from the use of the low contrast phantom, comprising 'lesions' of size well above system resolution but with contrasts ranging from greater than to less than the noise level, indicated that reading from the digital display gave better results than reading from film. It is believed that this gain in contrast perception resulted from the use of an appropriate window setting on the digital display such that noise was clearly visible.

2. INTRODUCTION

Work on digital systems at St Mary's Hospital, London, has largely focused on the evaluation of displayed image quality. The group has now had the opportunity to examine two different, commercially available PACS systems. This paper is concerned with the influence of a laser film digitizer on the system, the number of effective grey levels that can be created (i.e. the dynamic range of the system) and the perception of contrast in a noisy image.

3. MATERIAL

The system used for these tests comprised:

Two laser film digitizers, scanning spot size 210 microns with 180Mbyte local storage;

An optical disc archive, with a single disc drive (1Gbyte per side) and 3x825 Mbyte magnetic disc storage;

A display terminal consisting of 2 monitors with 180 Mbyte magnetic disc for local storage. Four 1280 line monitors with a band width corresponding to a 70MHz bandwidth and 8-bit video coding, non-interlaced.

4. METHOD

A set of six test films of different optical density ranging from 0.8 up to 2.5 was digitized; in addition a blank (filmless) image was also obtained. It was ensured that these test films were uniform and noise-free. Initial tests were performed using X-ray film exposed to give different densities, but these images were found to be very noisy and therefore unsuitable for testing the digitizer. Much of this noise was attributed to film processing. In order to avoid these problems the test images used here were generated from different thickness of **unexposed** film.

After digitization, the data was dumped on to an external computer, and mean pixel values and standard deviations within various sizes of regions of interest were computed, at various positions on the film.

5. THE LOW CONTRAST PHANTOM

In addition, a low contrast phantom was created by cutting holes in a series of superimposed blank sheets of film, to generate a set of holes of various sizes and contrasts, as illustrated in Fig 1. The sizes of the "lesions" ranged for $1cm^2$ down to $3mm^2$, much greater than the spatial resolution of the system. This phantom was then radiographed with various amounts of scatter, to create a set of 20 low contrast images. These were then digitized, and presented to observers who were requested to indicate how many lesions they could see on each image, both on film and on the digital display. They were free to adjust display window level and width. Since they knew the size and position of all the holes, this was akin to a 'ranking' experiment rather than a ROC study.

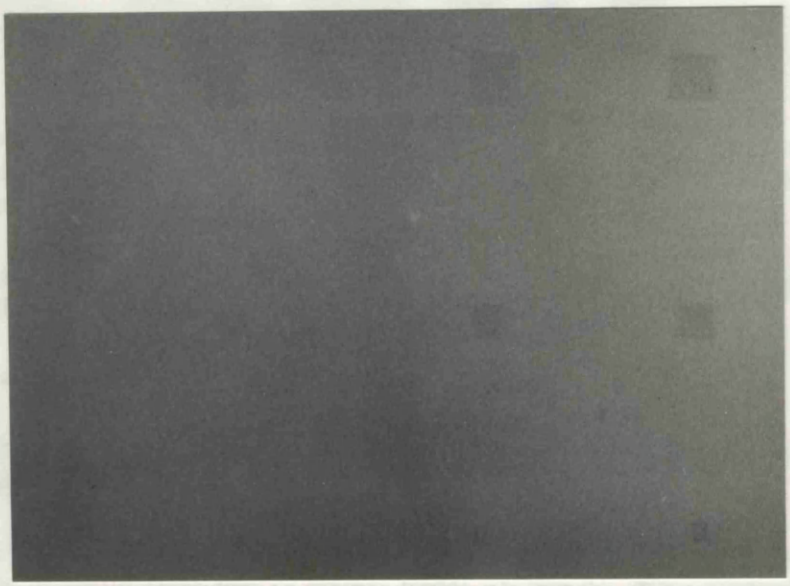


Fig 1. Contact print of part of film of low contrast test object.

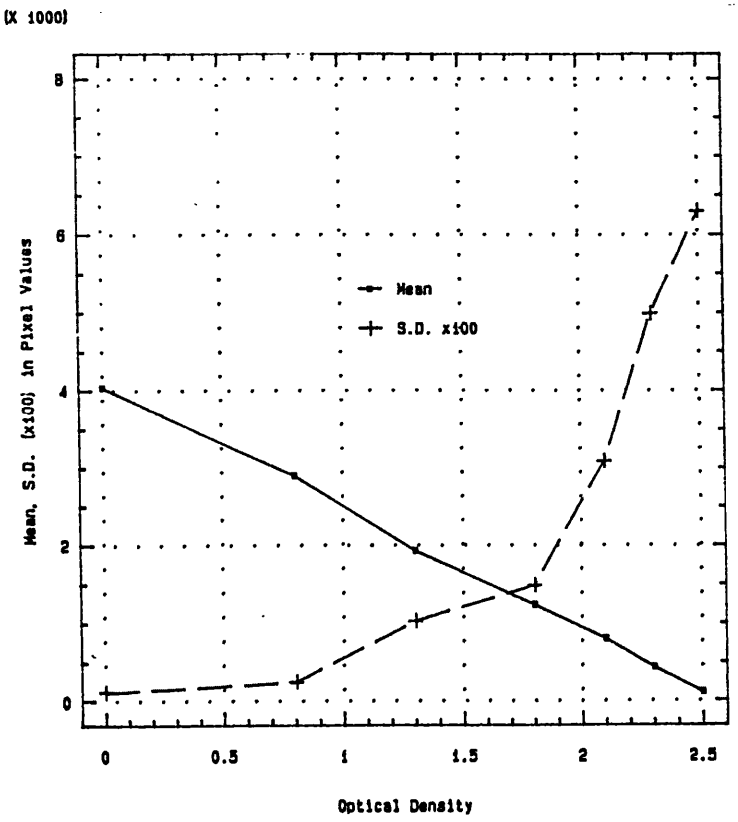


Fig 2. Mean pixel value and the corresponding standard deviation, plotted against optical density.

6. RESULTS

Fig 2 shows a graph of mean pixel value and the corresponding standard deviation for a central 256 pixel region of interest, plotted against optical density of the film for the uniform test film set. The standard deviation increases as the film density increases (and as the mean pixel value decreases).

From these values, an estimate of dynamic range R can be obtained. In order to compute the dynamic range R and obtain sensible values, it is necessary to invert the grey scale. R was computed as:

$$R = (Max-Mean)/S.D.$$

where Max is the Maximum pixel value (corresponding to the blank image), Mean and S.D. are the mean pixel values and their corresponding S.D. for a given optical density. This is plotted in Fig 3.

The results from the low contrast phantom indicated a significantly improved performance for the digital display, but did not appear to be critically dependent on the window levels that the observer chose. This is in keeping with the findings of others (for example Judy et al 1989). The histogram shown in Fig 4 plots the number of additional lesions seen either on film or on the digital display. This histogram is significantly ($p>0.001$) skew towards increased lesion detection on the digital display .

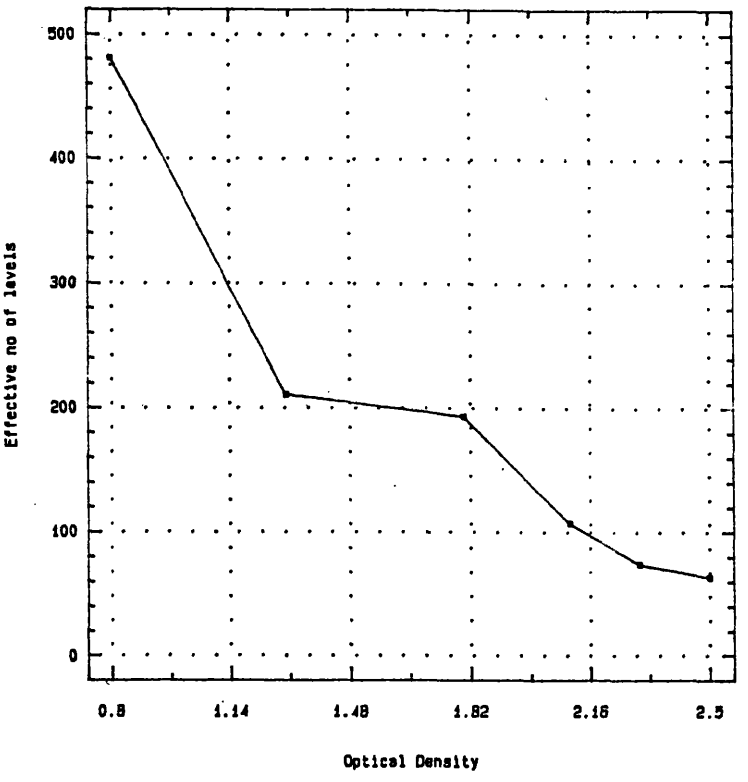


Fig 3. Dynamic range plotted against optical density.

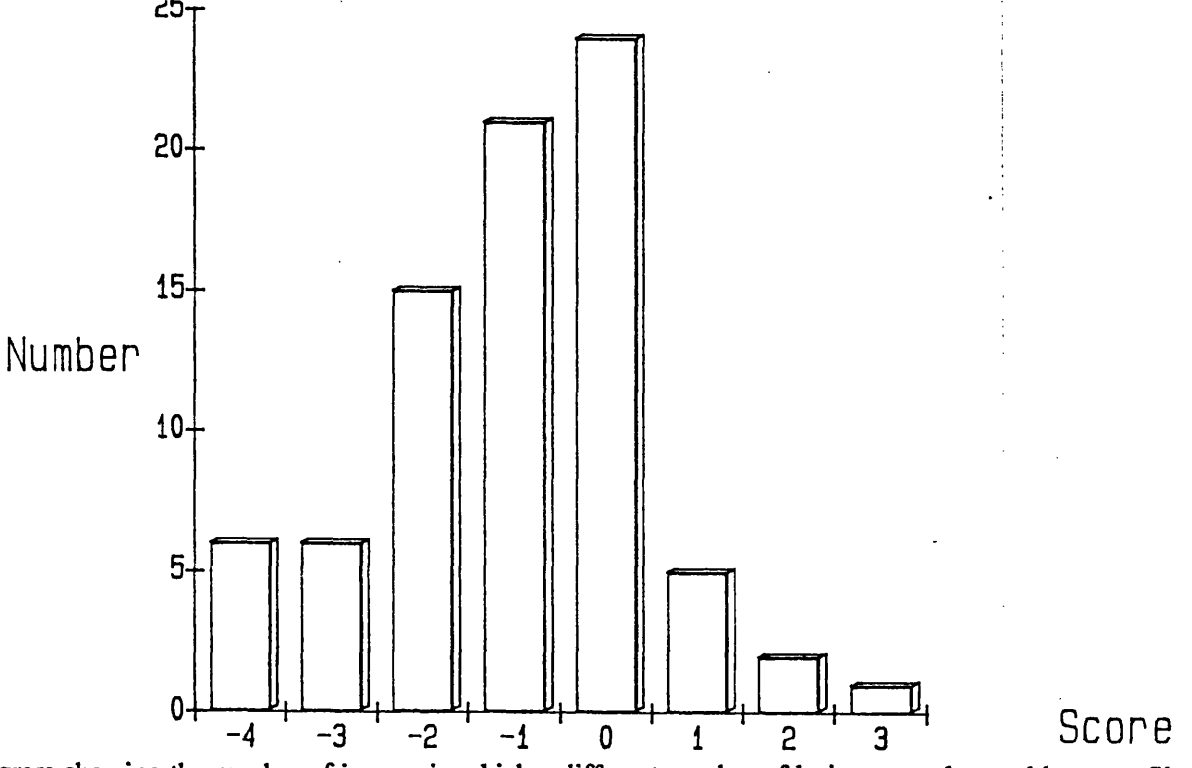


Fig 4. Histogram showing the number of images in which a different number of lesions was detected between film and the digital display. A score of zero indicates concordance, a negative score indicates that (that many) more lesions were observed on the digital display, and a positive score indicates that (that many) more lesions were observed on film. The difference between film and digital display is highly significant.

7. DISCUSSION

It may be observed from Fig 3 that the effective number of grey levels (the dynamic range) is less than 256 for most optical densities. Thus, although the data is being digitized to 12 bits, the least significant 4 bits contain mostly noise. This was for a 210μm spot, and it is anticipated that the noise level would rise with a decreased spot size. The signal to noise ratio is particularly poor for high optical density, when little light is transmitted.

On the other hand, contrast resolution with such a system is quite good. While some of the objects to be detected had much greater contrast than the noise level, the most difficult were concealed within the noise (as indicated by the estimates of noise presented above). Separate experiments have indicated that noise needs to be perceptible to the observer for reliable detection of such low contrast lesions. The perception of noise should be a weak function of the window setting chosen, and this is currently being testing with further ROC experiments.

8. CONCLUSION

Laser film digitizers are noisy, and it is hard to achieve the dynamic range desired for conventional projection radiography.

9. ACKNOWLEDGMENTS

We would like to thank the Department of Health, UK; Philips Medical Systems, UK; David Plummer of the Dept. of Medical Physics, UCL; and the Department of Medical Photography at St Mary's Hospital.

10. REFERENCES

1. P. F. Judy, R. G. Swensson, and P. Chopra, "Grading stenotic lesions: a perceptual study," Abstract Radiology, vol. 173(P), p225, 1989.

2. A. Todd-Pokropek, R. M. Dawood, J.O.M.C. Craig, H. Highman and H.I. Glass, "The evaluation of digitizer and video display systems in radiology: tests and test data sets," Oral presentation, EUROPACS 89, Brussels, 1989 (in press).

The optimal use of rating scales in R O C analysis

R M Dawood[1], A Todd-Pokropek[2],
J O M C Craig[1], J H Highman[1], A W Porter[1].

[1]Department of Radiology, St Mary's Hospital, London;
[2]Department of Medical Physics, University College London.

ABSTRACT

Observers participating in ROC studies are usually required to estimate the confidence with which each observation is made. With a discrete scale, the rating, or score, normally falls into one of 5 categories, ranging from 'definitely normal' to 'definitely abnormal'. However, a major problem in data analysis from ROC studies has been found to be caused by observers who have not used the rating scale in a uniform manner, and have made many responses corresponding to the two extreme categories with few responses falling in the middle. The use of a continuous rating scale, with a point selected using a mouse, has assisted in analysis, but only to a limited extent. It has therefore been suggested elsewhere that it is desirable to force observers to select intermediate points. The effect of such an approach on ROC curves was studied by asking a group of observers to re-score a set of difficult clinical images, after training and with continuous feedback on their compliance. Although the resulting fall in the ROC curves was not statistically significant, it is considered unwise to force observers to report in what to them appears to be an unnatural manner.

2. CLINICAL MATERIAL

A series of 40 well-validated hand images was selected for study. Half were from patients who had no known renal disease, and who had attended the X-ray department for a variety of unrelated conditions; the other half were from patients with proven chronic renal failure in whom a diagnosis of subperiosteal resorption had been evident radiologically on at least two occasions and in whom the diagnosis was also visible to the radiologist administering the tests.

The same image set has been used in previous ROC experiments[1,2]. The images were digitized from film, using a laser digitizer with a 210 μm spot size, and were displayed on a 1280 line monitor.

3. METHOD

The images were presented twice, on the same display system, to the same team of seven experienced observers, with a 10 month interval. On the first occasion, the observers were allowed to report the images as they pleased, with no special instructions regarding use of the rating scale. A continuous scale was used, and observers recorded their findings by selecting a point on the scale with a mouse (as illustrated in Fig 1). It was observed that, for this set of test images, a total of 56% of all reports fell into the two extreme categories (1.0 and 5.0). At the same time, a repeat trial for a limited set of images was conducted to confirm reproducibility.

After 10 months, the same images were re-reported, but the observers were instructed to attempt to use the rating scale in as uniform a manner as possible. To reinforce this request, a continuously updated histogram was displayed, showing how they had used the rating scale up to that point (see also Fig 1). The observers also knew that exactly half of the images were positive. The percentage of reports falling into the two extreme categories was reduced, but only to 32% of the total.

The ROC curves from the two studies were compared using the CLABROC continuous rating correlated data analysis program from the ROCFIT package as distributed by C.E. Metz.

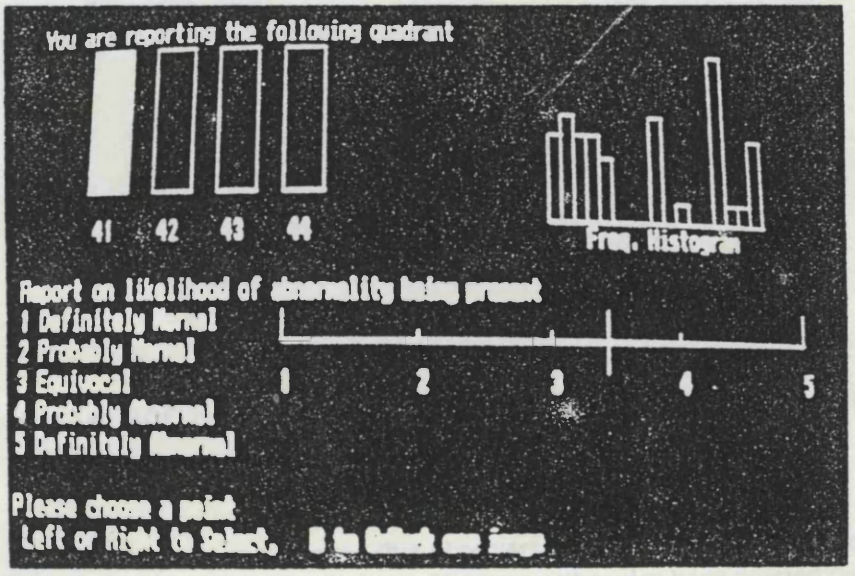


Fig 1. Data observation entry screen, showing continuously updated histogram.

4. RESULTS

The distribution of the points used on the rating scale during the two viewing sessions is shown in Fig 2. The corresponding ROC curves are plotted in Fig 3. The areas under the two curves were 0.8598 for the original data, and 0.8178 for the 'uniform' data, which were not significantly different (p=12%). In the original trial, the repeat experiments to check reproducibility showed much smaller differences than those observed here.

6. CONCLUSIONS

The re-training of observers to use the rating scale in a more uniform but artificial manner,

- can introduce bias,
- may lower resulting ROC curves,
- and should be avoided.

7. ACKNOWLEDGMENTS

We would like to thank our colleagues who acted as observers; the Department of Health, UK; Philips Medical Systems, UK; and the Department of Medical Photography at St Mary's Hospital.

8. REFERENCES

1. R.M. Dawood, J.O.M.C. Craig, J.H. Highman, J. Wadsworth, H.I. Glass, A.E. Todd-Pokropek, D.A. Cunningham, J.M. Stevens, A. Al-Kutoubi, R.W. Kerslake, A.H. Choudhri, C.J. Barber, M.E. Crofton and A.W. Porter, "Clinical diagnosis from digital displays: preliminary findings of the St Mary's evaluation project," Clinical Radiology, vol. 40, pp369-373, 1989.

2. R.M. Dawood, J.O.M.C. Craig, J.H. Highman, J. Wadsworth, H.I. Glass, A.E. Todd-Pokropek, D.A. Cunningham, J.M. Stevens, A. Al-Kutoubi, R.W. Kerslake, C.J. Barber, M.E. Crofton and A.W. Porter, "Evaluation of VDU displays in radiological diagnosis: a preliminary comparison of two systems," SPIE Proceedings: 1091 pp638-371, 1989.

3. C.E. Metz. ROCFIT software package, University of Chicago, 1987.

4. A.E. Todd-Pokropek, "Does pooling data change the validity of ROC assessments," Abstract Radiology, Vol. 173(P), p225, 1989.