THE ROLE OF PHYSICS TESTING IN BREAST CANCER SCREENING

A thesis submitted to the City University for the degree of Doctor of Philosophy

by Lynn Janette Martinez
Department of Systems Science

January 1995

Contents

Table of contents 2
List of tables and illustrations 7
Acknowledgements 10
Abstract 11
Glossary of terms used 12

1 Introduction to Breast Screening 16

1.0 Background to the UK breast screening programme 16

1.0.1 Objectives of this work 17
1.02 Outline of the study 17

1.1 Justification of breast screening 18

1.1.1 Rationale 18
1.1.2 Breast Cancer Prognosis and treatment 24
1.1.3 Cost/benefit and risk/benefit calculations 26

1.2 Screening Methods available 30

1.2.1 Breast self examination 30
1.2.2 Clinical breast examination 30
1.2.3 X-ray mammography 31
1.2.4 Xero-mammography 31
1.2.5 Ultrasound 31
1.2.6 Thermography 32
1.2.7 Transillumination 32
1.2.8 Magnetic resonance imaging 32
1.2.9 Digital mammography 33
1.2.10 Computed tomography 34
1.2.11 Genetic markers 34
1.2.12 Thermal bra 34

1.3 Survey of Current Screening programmes 36

1.3.1 USA 36
1.3.2 Australia 36

1.3.3 Iceland 36
1.3.4 Ireland 37
1.3.5 Holland 37
1.3.6 Sweden 37
1.3.7 UK 37
1.3.8 Finland 38
1.3.9 Canada 38

1.4 Strategies for implementing a screening programme 39

1.4.1 Risk factors 39
1.4.2 Target population 41
1.4.3 Screening interval 43
1.4.4 Call and recall 48

1.5 Treatment 49

2 The Role of Quality Procedures 51

2.1. Description of the breast screening service 51

2.2. Evaluating the decisions 56

2.3. Sensitivity analysis of changes in image quality 59

2.3.1. Mathematical model for changes in image quality 59
2.3.2. The results from the model 65

2.4. Desirable technical quality requirements of screening equipment 80

2.4.1. X-ray production 81
2.4.2. Image production by absorption in the breast 84
2.4.3. Production of the latent image on the film 85
2.4.4 .Processing the film 87
2.4.5. Reading of the film 87
2.4.6. Assessment equipment 88

2.5. Details of quality assessment 89

2.5.1. X-ray equipment 89

2.5.2. Processing 93
2.5.3. Viewers 95

2.6 Definitions of required quality - literature review 96

3 X-ray units 99

3.0 Introduction. Why the X-ray equipment is tested 99

3.1 Determining the optimum frequency for testing system components 99

3.2 Sources of noise and methods of determining their magnitude 100

3.2.1. Tube voltage 101
3.2.2. Output 107
3.2.3 Automatic exposure control 114
3.2.4. Half value layer 116
3.2.5. Image quality 118
3.2.6. Focal spot size 124
3.2.7. Alignment 125
3.2.8. Exposure time/ tube current 126
3.2.9. Compression 127
3.2.10 Radiation leakage 127
3.2.11. Field uniformity 128
3.2.12 Magnification 129
3.2.13. Grid factors 130

3.3. Optimum protocol determined from results 130

3.3.1. The size of variations leading up to tube failure 131

3.4. Alternative Strategy 134

3.5. Dose Monitoring 136

3.5.1 Requirements 136
3.5.2 Evaluation 137

3.6. Anode Material 138

3.6.1. Dose considerations 139
3.6.2. Image Quality 140
3.6.3. Screening statistics 141

4 Control of the film processor 143

4.1. Importance of the processor to image quality 143

4.1.1. The short term variations in batch processing 144
4.1.2. Medium Term Variations, between services 151
4.1.3. Long term variations, three years of sensitometry 158

4.2 Influence of the processor on X-ray images 162

4.2.1. Mean film density 162
4.2.2 Image quality 163

5 The Imaging Chain 171

5.1 Introduction 171

5.1.1. X-ray production 172
5.1.2. Image formation 172
5.1.3. Processing 173

5.2 Record keeping 175

5.3 Diagnostics 182

5.3.1 Block checks 182
5.3.2 X-ray sensitometry (stepwedge measurements) 182
5.3.3 Process control sensitometry 184

5.4 Fault detection 185

5.4.1 Processor 185
5.4.2 kV 186
5.4.3 Output of tube per mAs 188
5.4.4 Filtration 188
5.4.5 AEC 188

5.4.6 Screen sensitivity 189
5.4.7 Film variations 191
5.4.8 Focal spot 191

5.5 Financial analysis of new scheme 192

5.6 Monitoring daily measurements 198

5.6.1 Interpretation of the CUSUM graph 198
5.6.2 Computerised CUSUM 199

6 Conclusion 200

6.1 Value of Breast screening 200

6.2 Value of quality assurance 200

6.3 Optimising physical quality assurance 201

6.3.1 Financial benefits of quality assurance 202
6.3.2 Quality benefits of quality assurance 203
6.3.3 Organisational benefits of quality assurance 204

6.4 Further work 205

References 207

Bibliography 225

Appendix A Listing of the spreadsheet model of breast screening 249
Appendix B Screening statistics 260
Appendix C Sensitometry 262
Appendix D Calculation of contrast 265
Appendix E Statistical calculations 266

List of tables and illustrations

Chapter 1

Figure 1.1.1.1	Cumulative survival by size of cancer at detection	20
Figure 1.1.1.2	The size of screen detected cancers	22
Figure 1.1.1.3	The size of symptomatic cancers	22
Table 1.1.3	Cost of a range of healthcare options	27
Figure 1.1.3.1	Variations in cost per life year saved	29
Figure 1.4.1	Sensitivity of the breast to ionising radiation with age	40
Figure 1.4.2	Age-specific incidence rate and mortality rate for the UK	42
Table 1.4.3.1	The benefit of screening as a function of doubling time	44
Figure 1.4.3.2	Time gained per cancer due to screening as a function of screening interval	47

Chapter 2

Table 2.1	The variation in screening performance in different environments	54
Figure 2.1	Flow diagram of the screening process in NE Thames	55
Table 2.2.1	First round screening statistics	57
Table 2.2.2	Second round screening statistics to 31st March 1994	57
Table 2.2.3	Comparison of X-ray unit before and after a tube change	58
Figure 2.3.1.1	Illustration of how false positive and false negative decisions arise	60
Figure 2.3.1.2	True positives, false positives true negatives and false negatives	62
Figure 2.3.1.3	ROC curves for different values of sigma	64
Figure 2.3.2.1.1 to 4	Effect of incidence on false positives, cancer detection rate, false negatives and the number of cancers in the unscreened population	68, 69
Figure 2.3.2.2.1 to 4	Effect of non-attendance on false positives, cancer detection rate, false negatives and the number of cancers in the unscreened population	70,71
Figure 2.3.2.3.1 to 4	Effect of screening image quality on false positives, cancer detection rate, false negatives and the number of cancers in the unscreened population	72, 73
Figure 2.3.2.4.1 to 4	Effect of screening decision point on false positives, cancer detection rate, false negatives and the number of cancers in the unscreened population.	74, 75
Figure 2.3.2.5.1 to 4	Effect of assessment image quality on false positives, cancer detection rate, false negatives and the number of cancers in the unscreened population.	76, 77
Figure 2.3.2.6.1 to 4	Effect of assessment decision point on false positives, cancer detection rate, false negatives and the number of cancers in the unscreened population.	78, 79
Figure 2.4.1.1	Rotating anode tube	82
Figure 2.4.1.2	The X-ray spectrum produced by a Mo anode and 30 μm Mo filter	83
Figure 2.4.1.3	Relative exposure for constant signal to noise ratio	83
Table 2.4.3	The number of photons required for different contrast levels to be visible	86

Chapter 3

Figure 3.2.1.1	kV measurements from a system showing no drift	101
Figure 3.2.1.2	Warming up of the kV meter	102
Figure 3.2.1.3.1	Measured kV against distance of meter from the focal spot	103
Figure 3.2.1.3.2	Orientation of axes with respect to the X-ray field	104
Figure 3.2.1.3.3	kV meter function with position perpendicular to the anode cathode axis	104
Figure 3.2.1.3.4	kV meter function with position along the anode cathode axis	105
Figure 3.2.1.3.5	Measured kV as a function of the orientation of the meter	106
Figure 3.2.2.1	Central and East London mobile, output at 28kV	107
Figure 3.2.2.2	The warm up of a dosemeter	108
Figure 3.2.2.3.1	The position of the ion chamber within the beam	109
Figure 3.2.2.3.2	Inverse square law variations of the output	110
Figure 3.2.2.3.3	Output measurements with dosemeter position along anode cathode axis	111
Figure 3.2.2.3.4	Dosemeter response with position perpendicular to anode cathode axis	112
Figure 3.2.2.3.5	The effect of backscatter on output measurements	113
Figure 3.2.2.3.6	The effect of scatter from compression plate on air kerma	113
Table 3.2.3.2	Southend Basildon and Thurrock block check results, statistical summary	115
Figure 3.2.3.3	The effect of perspex position on AEC test	116
Figure 3.2.5.1	Image quality score as a function of optical density	118
Table 3.2.5.2.1	Summary of image quality score (calcifications included)	120
Table 3.2.5.2.2	Summary of image quality score (calcifications excluded)	120
Figure 3.2.5.2.1	Mean score of ten films by day	121
Figure 3.2.5.2.2	Mean score of ten films by day calcifications excluded	121
Figure 3.2.5.2.3	Score due to small calcifications	122
Figure 3.2.5.2.4	Mean score for each film	123
Figure 3.2.5.2.5	Film score as a function of optical density	123
Figure 3.2.5.3.1	Image quality score as a function of optical density	124
Figure 3.2.5.3.2	Image quality score as a function of tube voltage	124
Table 3.3.1.1	Change of QA parameters with time SBT mobile	132
Figure 3.3.1.1	Output of SBT mobile as a function of time	132
Figure 3.3.1.2	Output coefficient at 28kV CEL and SBT mobile	133
Figure 3.4.1	mAs as a function of time Southend Basildon and Thurrock	135
Table 3.4.1	The drift in mAs measured - Southend mobile	135
Table 3.5.2	Mean compressed breast thickness	138
Figure 3.6.1.1	Distribution of dose in NE Thames BSP	139
Figure 3.6.1.2	Distribution of dose normalised to an optical density of 1.4	140
Figure 3.6.2.1	Distribution of image quality scores in NE Thames BSP	140
Table 3.6.3.1	Screening statistics Southend and BHB	141
Table 3.6.3.2	Assessment statistics Southend and BHB	142

Chapter 4

Table 4.1.1	Drift in sensitometry parameters during batch processing	146
Figure 4.1.1.1	Variation in OD following batch exposures and processing	147
Figure 4.1.1.2	Variation in mAs following batch exposures	148
Figure 4.1.1.3	OD as a function of mAs for batch block checks	149
Figure 4.1.1.4	Fog, Dmax and gradient during batch processing	150
Figure 4.1.1.5	Speed during batch processing	150
Figure 4.1.2.1 to 4	Fog, speed gradient and Dmax data for six months at Whipps Cross	152,153
able 4.1.2	Dates of chemicals being replaced at Whipps cross	154
Figure 4.1.2.5	IQ score as a function of optical density	155
Figure 4.1.3.1 to 4	Fog, speed, gradient and Dmax data for three years at Whipps Cross	159,160
Table 4.1.3.1	Process control statistics, speed at Whipps cross	161
Table 4.1.3.2	Coefficient of variation of processor parameters	162
Figure 4.2.1.1 to 4	Block check OD as a function of PC parameters	164 et seq.
Table 4.2.2	Process control parameters and image quality scores from physics checks	168
Figure 4.2.2.1 to 4	Fog, gradient, speed and Dmax as a function of image quality	168,169,170

Chapter 5

Figure 5.1	Ishikawa diagram of imaging chain	174
Table 5.2.1	Sources of faults and identifying tests	181
Figure 5.4.2	Stepwedge attenuation	187
Figure 5.4.5	Relationship between contrast index and characteristic curve	189
Table 5.5.1	Comparative costs of QA regime	194
Table 5.5.2	Capital costs	195
Table 5.5.3	Cost of daily checks	196
Table 5.5.4	Physics checks	197

Chapter 6

Table 6.2	Summary of the effect of changing image quality and decision point on the number of cancers detected	201

Acknowledgements

I would like to express my thanks to everyone involved with the breast screening programme in North East Thames, I have been given a warm welcome wherever I have been working and have been provided with data for this project by many people. Two people warrant a special mention, Susannah Badger, who has explained the subtleties of the statistical returns to me and Carol Turner who has faithfully sent me her QA and sensitometry films every month for three years.

I would like to thank my supervisors, Dr Fred Charlwood for his encouragement and unceasing patience through endless redrafts of this document and Professor Denis Hill who set the ball rolling.

I am very grateful to my good friends Rupert Francis and Andy Green both of whom have helped me to sharpen my ideas during numerous discussions and have helped me practically, Rupert by collecting data when I have been ill and Andy by allowing me access to his computing facilities.

Last but not least, many thanks to my husband Joe who has kept me going with words of encouragement, prayers, food and cups of tea, and who has been a very thorough proof-reader.

Declaration

Abstract

The aim of the work contained in this thesis is to critically evaluate the role of quality assurance testing on equipment used within the breast screening programme in the UK. At the time the work began, mass screening had only just started in the UK, and, although many countries had breast screening projects of one form or another, no-one else was attempting to screen the whole population. The X-ray equipment was totally new, having been re-designed in line with the specifications from the Department of Heath [1], and it was far from certain that the quality assurance procedures recommended were the most appropriate or the most cost-effective for this particular branch of imaging.

Chapter one discusses the evidence and rationale for doing breast screening at all. The range of available screening techniques is described and their potential benefits for breast screening are discussed. Strategies for screening are also examined and a mathematical model to relate the benefit of screening to rate of cancer growth and screening interval has been developed. Breast screening programmes both in the UK and abroad are reviewed.

In chapter two a statistical description of screening in North East Thames is presented and using real statistics as the starting point, a computer model has been developed which uses three levels of Bayesian likelihood analysis to represent the screening, assessment and biopsy stages. Sensitivity of the cancer detection rate and number of missed cancers to variations in uptake, image quality and decision criteria is analysed particularly to show how important image quality is to the final outcome of the screening process. The procedures for physical quality assurance are described.

Chapter three contains analysis of the gathered data from X-ray equipment and finds that the X-ray tube output is a key indicator of tube condition. The minimum period of time at which such changes are detectable is calculated.

Chapter four examines the role of the film processor and analyses the key sources of variation in processing.

Chapter five takes the results from the previous two chapters and uses them to build a new scheme of quality assurance which provides more information and better analysis for less effort. A financial analysis has been done comparing the new and old systems.

Chapter six concludes that breast screening QA can improve the effectiveness of screening and concludes that the scheme developed in this work enables it to be done more cheaply and efficiently. Areas which are unresolved by this project are identified and further work is suggested.

Glossary of terms used

Accuracy the proportion of correct results out of all results.

AEC (Automatic Exposure Control) A mechanism used to terminate an exposure when the appropriate amount of radiation has been produced.

AED (Automatic Exposure Device) A mechanism used to terminate an exposure when the appropriate amount of radiation has been produced.

Anode the positive terminal of an electrical device (in this document, the X-ray tube) used synonymously with the term *target*, although strictly speaking, the target is a part of the anode assembly.

Artefacts features on images which do not represent the imaged tissue, they result from faults in either the equipment or the film processing.

Beam Hardening The process whereby the X-ray beam is differentially attenuated as it passes through matter, the low energy photons are preferentially absorbed leaving a beam which has a much higher proportion of high energy photons, known as a hard beam.

Cassette light tight box which houses an intensifying screen and the X-ray film during exposure an prior to development.

Cathode The negative terminal of an electrical device - synonymous with *electron gun* in this document

Characteristic curve A plot of the film response (optical density) to exposure. This characterises the way in which the film and intensifying screen convert the subject image formed by variations in X-ray distribution into a photographic image

Contrast agent A biologically compatible material of high attenuation which is used to enhance attenuation differences in the body.

CUSUM the CUmulative SUM of the deviation of test results from the target. This is a sensitive mathematical technique to detect drift in test results.

CV coefficient of variation. Statistical measure of the variability of a set of results equal to the standard deviation divided by the mean. Often expressed as a percentage.

Exposure (1) an amount of radiation

Exposure (2) the act of producing radiation

FFD The focus to film distance

FHSA Family health services authority, the organisation which is responsible for co-ordinating the activities of GPs and for keeping patient records.

Filter thin layer of metal through which a proportion of the X-ray photons can pass.

Gassy An X-ray tube which no longer has a good vacuum within the glass envelope is said to be gassy, this can give rise to arcing during an exposure.

HVL half value layer, a measure of the penetrating power or 'hardness' of an X-ray beam. The thickness of material which is required to reduce the intensity of the beam by a factor of two.

Intensifying Screen a fluorescent layer used to convert X-rays to light, usually located within a cassette.

K-edge This refers to the energy required to excite an electron in the innermost shell (k shell) of an atom ionisation. The X-rays produced by the filling of the K-shell have the same energy which is characteristic for each element.

K-edge filter A material which is placed in the radiation beam in order to alter the spectral distribution of photon energies. A k-edge filter is one in which the k-edge occurs mid way through the spectrum thus enhancing the number of photons in the spectrum at the k energy and strongly attenuating just above the k energy.

kV Kilo volts. Thousands of volts. In this application it is synonymous with the voltage across the X-ray tube.

Magazine A light tight box which is used for storing undeveloped films. Used in daylight processors and when no processing facilities are available locally for example on a mobile mammography unit.

mAs (milli ampere second) - the product of the current flowing through the X-ray tube and the time for which it flows. Equivalent to the electrical charge striking the anode.

NPV negative predictive value - the proportion of true results out of all positive results.

OD optical density, a logarithmic measure of the blackening on a film.

PNL Prior notification list. This is a list of names and addresses which are sent to the GP to be checked for inaccuracies and their suitability for screening

PPV positive predictive value - the proportion of true positive results out of all positive results.

Push-processing Deliberately overdeveloping a film in order to increase its photographic speed.

QARC quality assurance reference centre

ROC analysis Receiver Operating Curve analysis. A method of evaluating image quality which involves decision making based on a large number of randomly presented radiographs. The observer is expected to classify the radiograph as positive, ie object present or negative ie object absent. The curve is a plot of the proportion of true positive results against the proportion of false positive results.

Screening examining every member of a supposedly healthy population for a disease which has not yet manifested itself.

Sensitometry a process which measures the sensitivity of a film to a range of exposures.

Sensitivity the proportion of true positive results out of all positive conditions.

Signal to noise ratio A measurement of the detectability of a feature in the image. When close to quantum limits, the signal can be considered to be equal to the

number of photons in the image feature, n, and the noise the square root of this number, $\sqrt{n}$, as the number of photons increases the signal to noise ratio also increases making the image feature easier to detect.

Specificity the proportion of true negative results out of all negative conditions .

Chapter 1

Introduction to Breast Screening

1.0 Background to the UK breast screening programme

In 1986, the government of the United Kingdom set up a working party headed by Professor Sir Patrick Forrest to look into the possibility of doing population screening for breast cancer in the UK. The conclusive recommendations of the working party are contained in the Forrest report [2]. In the period between 1987 and 1990, a nationwide network of breast screening units was established in line with the recommendations of the report. Included in the recommendations were very firm guidelines on the necessity of additional training for all of the involved professions and strict procedures for quality assurance.This aspect was further emphasised by the publication of the Pritchard report [3] which quantified performance targets for the breast screening service. As a consequence, breast screening within the National Health Service had a quality assurance network included in its funding from the outset. Each Regional Health Authority (or in the case of Wales, Scotland and Northern Ireland, Health Board) was required to establish a Quality Assurance Reference Centre (QARC) which was to be responsible for collection of data on the standards being achieved, and provision of advice to the screening and assessment units for which it was responsible. To ensure that there was uniformity across the country and that "best practice" was widely disseminated, the breast screening service set up committees for each discipline, comprising a representative from each region in England, a representative from Scotland, Wales and Northern Ireland, and a representative from the private sector. These are known as "Big 18s" as the membership comprised 18 representatives plus the National Co-ordinator, and provided a means of direct communication between regions which was not normally available to service staff. These structures and the staff required to perform quality monitoring cost a considerable amount of money, some 10% of the running cost of the programme as a whole (excluding capital costs), and therefore the performance of the quality assurance network also needs to be placed under scrutiny. The work described in this thesis looks critically at one branch of the quality assurance network, that of equipment testing. This job is performed primarily by a physicist on a periodic basis, but is also complemented by daily measurements performed by the radiography staff on site. When the working party from the IPSM produced a document containing test methods [4] and recommended performance parameters of equipment, a great deal was

borrowed from well-established methodology used in other areas of diagnostic X-ray work, and the performance parameters were "best-guesses" based on experience of other X-ray equipment. At the time however, very few people had much experience of mammography, and even fewer of quality assurance within mammography

1.0.1 Objectives of this work

This piece of work aims:

1. To identify the parameters which are most likely to give some warning of tube failure.

2. To assess the most cost-effective way of monitoring the performance of equipment without compromising on quality standards.

3. To establish procedures which allow cost-effective monitoring to be performed and timely interventions to be made

4. To define the minimum data set required to be confident that equipment is performing satisfactorily.

1.0.2 Outline of the study

First, the need for breast screening will be considered, then, using the statistical data gathered from the programme within North East Thames, the need for physical quality assurance within the programme will be evaluated. The results from routine tests will be analysed to see which particular tests are of value and where quality assurance efforts should be concentrated and finally proposals for an improved system of quality assurance will be made.

Data has been collected over a period of three years in North East Thames Regional Health Authority within the breast screening programme. The measurements include:

1 Those quality assurance measurements which are required periodically on X-ray and associated equipment and are performed by a physicist

2 The routine checks performed by the radiographers on the X-ray machine and on the processor

3 The statistical data which is held on computer for the whole of the regional breast screening programme

4 Results from shorter controlled experiments to examine specific issues such as the effect of variations in set up when measurements are made.

1.1 Justification of Breast Screening

Like any other health activity, breast screening must be evaluated to see if it capable of achieving and actually achieving what it set out to. If not, the activity cannot be justified. Health care funding is overstretched and needs to be targeted where it can produce the best result.

1.1.1 Rationale

Breast cancer kills more women in the United Kingdom than any other type of cancer; 15,000 deaths per annum, which represents 20% of all cancer deaths in the female population, as approximately 50% of deaths in women are due to cancer of some sort [5], breast cancer accounts for 10% of all deaths in the female population. Kalache [6] has combined data from many sources to give a world-wide picture of the effect of breast cancer; in the UK, the incidence is low compared to most developed countries, however, the mortality is one of the highest. A number of possible explanations suggest themselves: treatment might not be as effective as that given elsewhere, the cancers may be more aggressive or the cancers may be going undetected for longer. There is no reason to suppose that the cancers in the UK are particularly aggressive and with modern communications, it is unlikely (although it cannot be ruled out) that treatment is any different from the rest of the world. The prognosis for breast cancer improves greatly if the cancer is treated at an early stage. There are a variety of tests available which are able (to varying degrees) to detect breast cancer, it should therefore be possible to detect breast disease at an early stage, thereby increasing the survival rate and bringing down mortality due to breast cancer (in a specific population).

The World Health Authority [7] has set criteria which should be met if a population screening programme is to be implemented.

i The disease in question should pose a significant health problem.

Breast cancer certainly poses a significant health problem accounting for approximately 10% of all female deaths in the UK and, because it usually occurs earlier on in life than heart disease, which is the most common killer for women of all ages, it accounts for an even greater proportion of life years lost [5].

ii The natural history of the disease should be understood.

The natural history of breast cancer has been well studied at the clinical phase to death (whether by cancer or other cause) but in the pre-clinical stage there is little data on human breast cancer; information is generally based on animal studies.

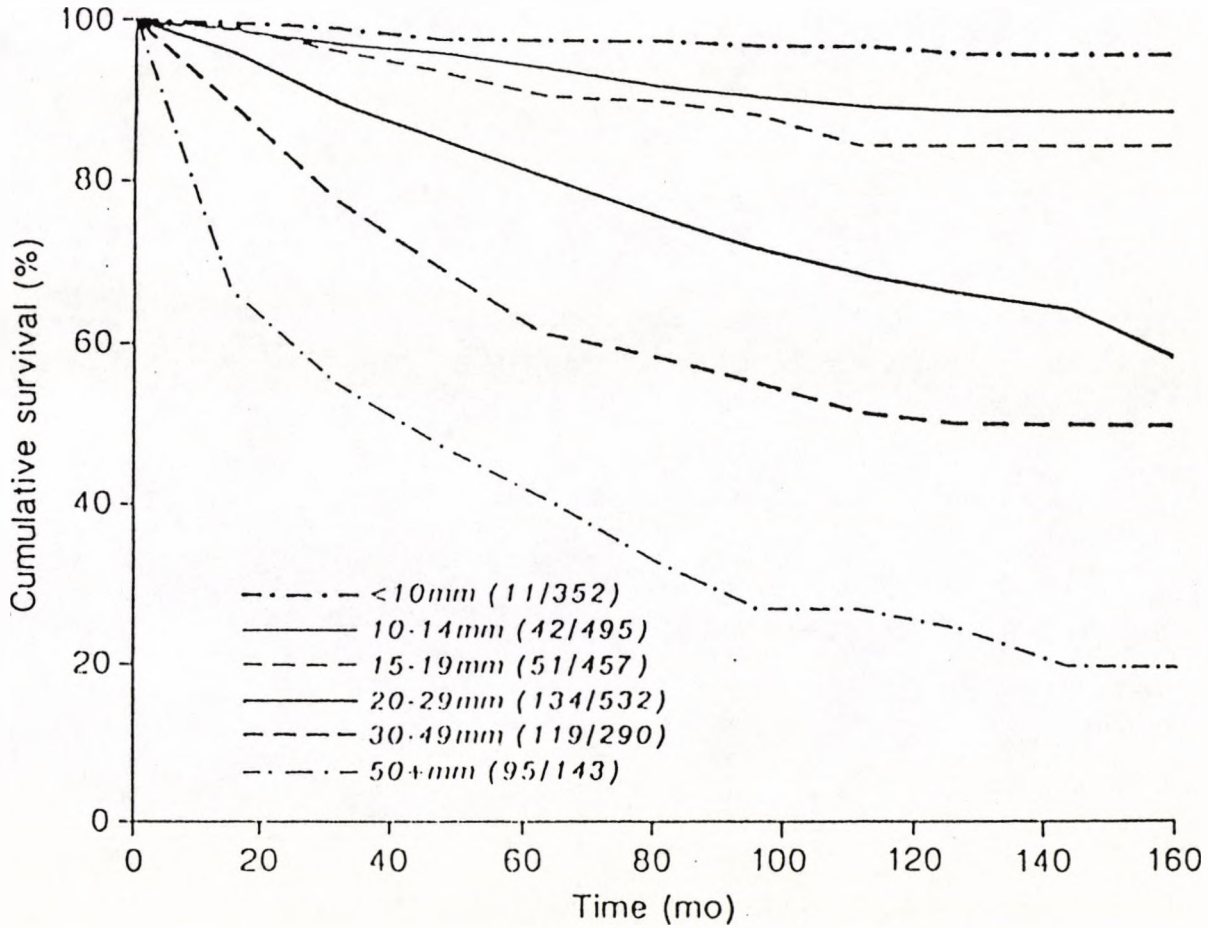
iii There should be a recognisable early stage of the disease and treatment at this stage should be more beneficial than treatment later on.

The early stage of the disease is a small cancer which has not yet metastasised. There is considerable evidence that cancers which have not yet begun to spread are much more likely to respond well to treatment [8,9,10,11,12]. There is also evidence that smaller cancers are likely to have a better survival rate than large ones; this suggests that small cancers are likely to be those which have not yet metastasised. This is illustrated in figure 1.1.1.1 taken from Tabar [13].

iv There should be a suitable test available which is acceptable to the population.

There are a number of tests available, including breast self-examination, X-ray mammography, Xero-mammography and ultrasound. Currently X-ray mammography is the screening technique of choice, the reasons for this will be discussed later, and various studies have achieved an uptake of between 50% and 90% proving that it is well accepted by most women.

Figure 1.1.1.1 Cumulative survival by size



(from L.Tabar, reference [13])

The question must be asked whether mammographic screening does in fact do what it claims and detect breast cancer at an earlier stage. Data from Dr Clive Wells at St Bartholomew's Hospital demonstrates that cancers detected by the screening programme are, on average, smaller than those detected by the women themselves; this is illustrated in figures 1.1.1.2 and 1.1.1.3

v *There must be adequate facilities available for the diagnosis and treatment of the abnormalities detected.*

In the UK, the treatment facilities already existed for a disease which has an exceptionally high mortality rate in this country [6]. It was expected that the first round of screening would produce a sharp increase in workload which would then return to its normal level once a steady state has been reached where the cancers detected by screening correspond to the incidence of breast cancer. The provision of diagnostic facilities is implicit in the setting up of a screening programme. The first round increase in treatment workload could have been considerably eased by phasing in screening by for example starting with a small age range of women and building up to cover the whole age range over a period of several years.

vi *If the onset of the disease is insidious, screening should be repeated at regular intervals the frequency of which depends on the natural history of the disease.*

The screening interval is one of the most difficult parameters to decide. The UK has the longest screening interval of any of the currently running programmes. Analysis of the screening data as it becomes available should enable the screening interval to be adjusted if necessary. This is discussed further later on in this chapter.

vii *The chance of harm should be less than the chance of benefit.*

The risk of physical harm (inducing a cancer) is very much smaller than the likelihood of physical benefit (increasing survival by early detection). Data from screening trials summarised by Shapiro [14] indicates that this is usually the case in the over 50 age group.

FIGURE 1.1.1.2 Size of cancers detected by screening

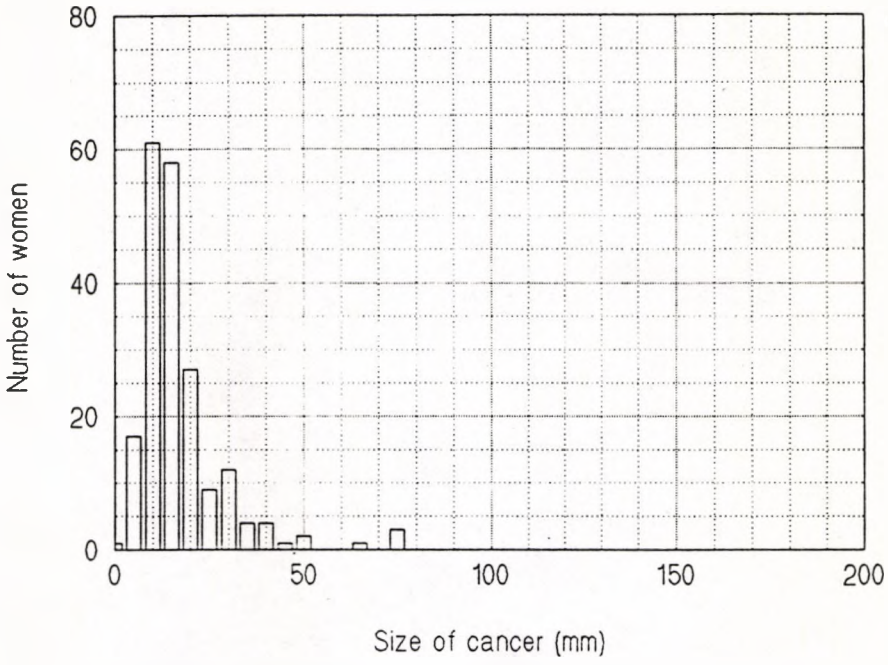
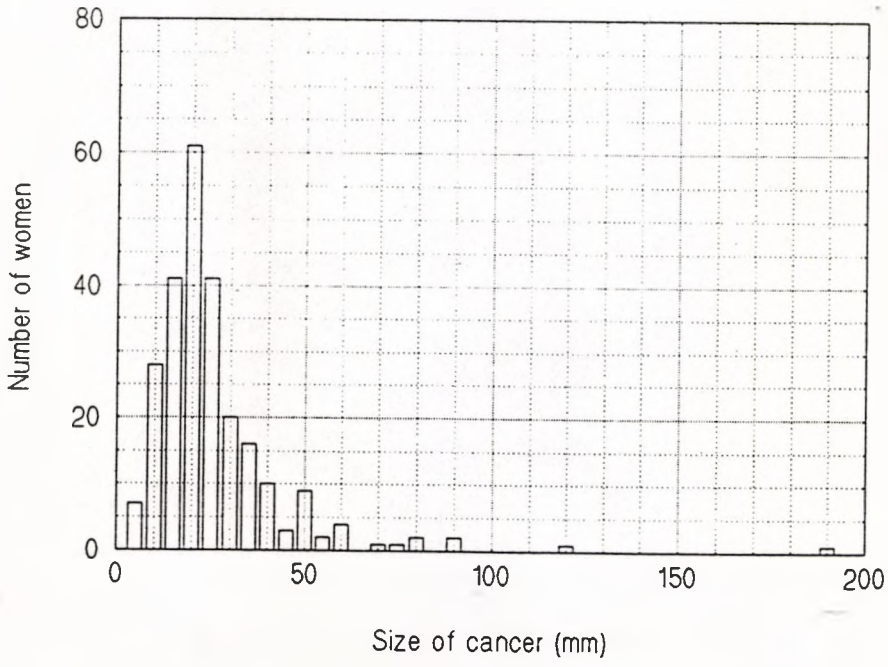


FIGURE 1.1.1.3 Size of cancers found symptomatically



Data courtesy of Dr C Wells, St Bartholomew's Hospital

The potential benefit to a woman engaging in screening is that a breast cancer which she cannot detect herself may be revealed by mammography thus enabling her to undergo less radical or more effective treatment because the disease is still in an early stage.

The potential of physical harm comes from two sources,

a The induction of breast cancer [15,16]
b Having to undergo diagnostic investigations such as fine needle aspiration, biopsy or treatment when there is no cancer present.

Psychological harm must also be considered. Medical examinations in general are stressful, and more so examinations for life-threatening diseases [17]. Being invited for mammography may in itself cause distress and being called back for second level assessment will cause greater stress, even when the final outcome proves negative.

viii *The cost of the screening programme should be balanced against the benefit.*

The cost calculations are very complex but generally show that breast screening is an activity which has similar cost-benefit values to other widely accepted medical practices [18].

Overall, in the Forrest report [2] it was considered that benefits do outweigh the risks. This was a controversial decision considered by some people to be a political vote-catcher prior to an election; others argue that high financial costs and dubious benefits make breast screening a much less worthwhile exercise than anticipated [19,20].

1.1.2 Breast Cancer Prognosis and Treatment

There is a fundamental underlying assumption in screening that early detection of the disease improves the prognosis. In order to define exactly what is meant by "early", some medical terminology is necessary. The classification of the development of a disease when it has been diagnosed is known as staging. Staging of breast cancer is done in the following way [21].

Stage 0 in situ disease

Stage I lesion < 2cm, no spreading of disease

Stage IIA no evidence of tumour but movable axillary lymph nodes involved
OR lesion < 2cm plus movable axillary lymph node involvement
OR lesion between 2 and 5 cm but no lymph node involvement

Stage IIB lesion between 2 and 5cm and movable axillary lymph node involvement
OR lesion greater than 5cm but no lymph node involvement

Stage IIIA any tumour with fixed axillary lymph node involvement

Stage IIIB tumour spread to chest wall or skin plus any lymph node involvement
OR any tumour plus mammary chain lymph node involvement

Stage IV metastatic disease

The prognosis for breast cancer has been shown to depend strongly on the tumour size at diagnosis [22] and also upon the nodal involvement. Gusterson [11] quotes 20% to 30% relapse rates for women with no nodal involvement and attributes at least some of the relapses as incorrect staging due to poor sampling of the nodes. The American College of Surgeons did a survey [22] across the country looking at the prognosis of breast cancer and found that cancers with no nodal involvement had a significantly better prognosis than those with nodal involvement and also that the smaller the cancer, the better the prognosis, although this may be due to smaller cancers being much less likely to have any nodal involvement. They also found that cancers which occurred in the lateral half of the breast had a better prognosis than those in the medial half. Black women appeared to have a much worse prognosis than their white counterparts when the data was studied. This could be accounted for

entirely by the fact that on average, black women were diagnosed as having cancer at a much later stage; low standards of healthcare within the black community are likely to be the culprit for this. There does not appear to be any intrinsic racial factor determining the prognosis of breast cancer.

Different types of cancer also have very different prognoses. For example, ductal carcinoma in situ (DCIS) has an excellent prognosis with only 10% (Fentiman [23]) of women going on to develop invasive carcinoma when given no treatment, but there is some disagreement on this figure, Hayward [12] quotes 50% do not become invasive; for the same condition, total mastectomy has a 99% cure rate.

Other indicators of outcome are the presence of a variety of oncogenes associated with metastatic potential [24]. If the risk of metastases is high, adjuvant systemic therapy (chemotherapy or tamoxifen) is likely to be of benefit.

The prognosis is greatly improved for non-metastatic disease and for small tumours, the expectancy will depend on the chosen treatment regime. This will be discussed in section 1.5.

Rubens [25] calculates that the routine use of adjuvant systemic therapy has increased the ten year survival by 10%, with the most benefit shown in the tumours with poor prognosis (stages III and IV). Which would account for 2,000 fewer cancer deaths in the UK.

Spittle [26] compared the traditional complete mastectomy with the more recent regime of lumpectomy plus radiotherapy and found that the survival was the same for both groups but that lumpectomy was more likely to have a local recurrence.

Joensuu [27] found that when women who had been screened were compared with women who had self-diagnosed cancer, there were no differences in the size and type of tumours, but that other factors which influence the outcome, such as the presence of metastases, are more favourable in the screened group.

1.1.3 Cost-Benefit and Risk-Benefit Calculations

Evaluating the cost-benefit of screening is a highly contentious area because of disagreements about which items should be included in the analysis and how exactly the analysis should be done. A very simplistic way of calculating the cost benefit is to make an estimate of the number of lives saved by the programme, and then divide by the total cost of the programme. Using data from the Forrest report [2] gives a figure for single view mammography of £12.86 per screen. 0.0049 cancers are detected per woman screened which equals £2624 per cancer found (1986 prices). Only 62.5 % of those cancers found would have resulted in death and approximately one third of these will benefit by early diagnosis, therefore we have to divide the figure per cancer found by 0.625 and then by 1/3 to account for the expected mortality reduction; thus we reach £25,190 per life saved.

The cost per screening examination depends very much on the accounting practices used, such as the method of writing off capital equipment costs and the discount rate used, and whether one includes training and research in the equation or not. It must be recognised that in an area as politically sensitive as this, there will be cost estimates varying by an order of magnitude depending on whether the instigator wishes to proceed with screening or wishes to oppose it.

Health economists prefer to work in terms of cost per life year gained or to be even more sophisticated, cost per quality adjusted life year (QALY) gained, the assumption being that a year of life in perfect health is worth far more than a year of life in very poor health. If a uniform method of doing the calculations can be found, judgements on how to spend Health Service resources are then made much simpler.

This has only taken account of the screening programme itself. Within the Health Service it is wise to look at the wider implications since funding for all parts of the Health Service comes from central government. First, the treatment cost must be considered. If there is no over diagnosis i.e. cancers diagnosed which are not really there, or, which would not have caused a health problem, then once the programme is in full swing, the number of cancers found through the programme will not be greater than would have occurred anyway; they will simply be detected earlier on in their development. This ought to mean that treatment costs will be

reduced because cancers caught earlier are easier and cheaper to treat. This saving should then be offset against the cost of the screening programme. The Dutch analysis [28] estimated that once a steady state situation had been achieved, 45% of the cost of the screening programme would be found from the savings in treatment costs. Taking this into account, the cost per life saved under the scheme proposed in Forrest [2] is £13,854.

From a government point of view, one must also consider the cost to the economy due to time taken off work to attend screening offset against the losses or gains in time due to easier treatments and the consequences in earnings and tax revenue.

When van der Maas tried to do this for the government in Holland [28] it was found that of all the variables, the discount rate (to adjust for inflation for expenses or savings which occur later on in the programme) was the factor which had the greatest effect on the costs calculated. He also concluded that mass screening yields diminishing returns the more intensive it becomes. Mushlin and Fintor [29] in the USA looked at nine computer based cost models from different research centres which gave cost per life year saved ranging from $3,400 to $46,600 depending on the screening regime and the way various quantities were calculated. They conclude that although not the best buy possible in healthcare, breast screening was within the acceptable range of healthcare procedures, costs for such procedures are in table 1.1.3.

Procedure	$ per life year (1991)
Coronary artery bypass graft, left main	7,300
Mild high blood pressure	32,600
Breast cancer screening	20,000 - 50,000
Liver transplant	225,000
Low osmolar contrast agents	228,000
Coronary artery bypass graft, angina	62,900

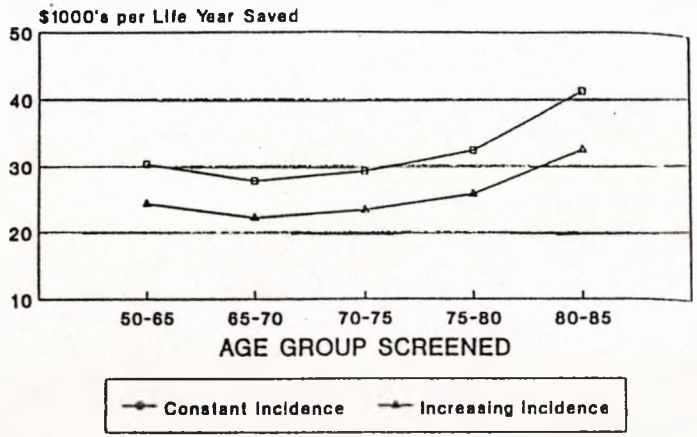
Table 1.1.3 Costs of a range of healthcare options, data from reference [29]

Brown [18] did a sensitivity analysis of the cost per life year saved as a function of the screening interval, age, price of screening examination,

false positive rate and incidence. These graphs are reproduced in figure 1.1.3.1. As might be expected, as the incidence rises, the cost per life year saved becomes less, as more cancers will be detected for any given screening effort. False positives are obviously costly, and show a linear relationship with cost per life year saved. As the screening interval goes up, the associated cost goes down due to the decreased frequency of screening but not in direct proportion to the reduction in workload. This is because as the screening interval increases, a greater number of cancers become clinically manifest in between screens and these cancers obtain no benefit due to screening. Finally, the cost per examination also increases the cost per life year saved in direct proportion.

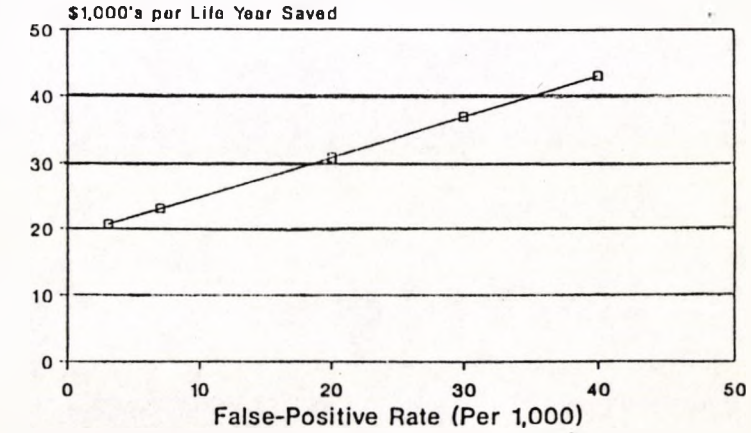
There are adverse effects due to a screening programme most importantly the risk of radiation induced cancer. Skrabanek [30] discusses other adverse effects including causing anxiety to the screened population and the possibility of over diagnosis and associated unnecessary medical intervention.

Figure 1.1.3.1 graphs reproduced from reference [18] showing the variation in costper life year saved.



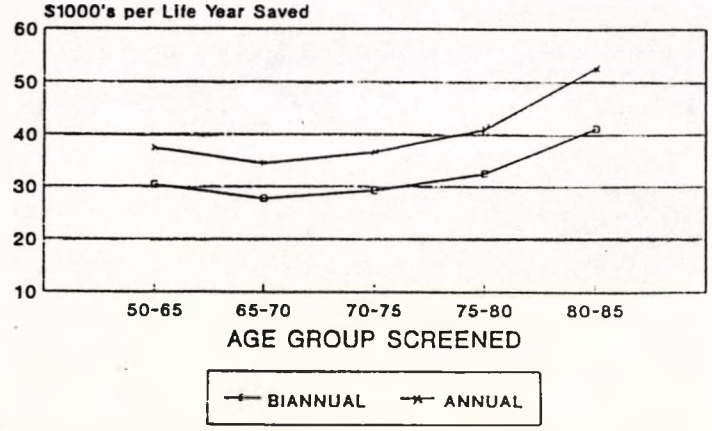
Biannual screening; Price = $55; ($20 for CBE alone); Increasing incidence = 1.2% per year

Breast cancer screening cost-effectiveness by incidence.



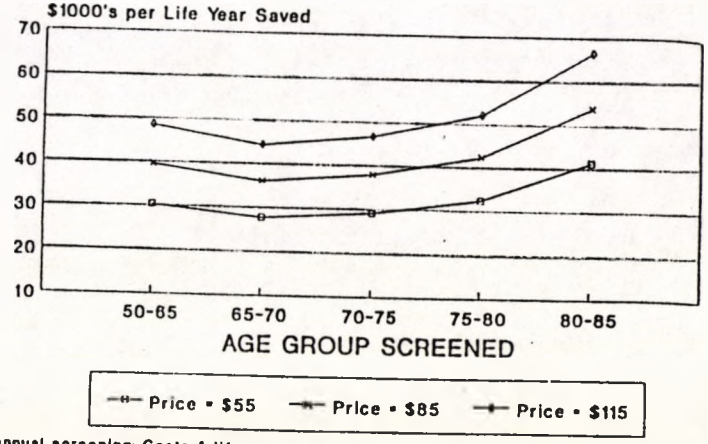
Biannual screening, Price = $55
Price of alternating CBE = $20
Life years and costs discounted at 5%

Breast cancer screening cost-effectiveness by false-positive rate.



Price = $55; Price of alternating year clinical breast exams = $20; Life years and costs discounted at 5%

Breast cancer screening cost-effectiveness by screening interval.



Biannual screening; Costs & life years discounted at 5%; Price of alternating year clinical breast exams = $20

Breast cancer screening cost-effectiveness by age and price

1.2 Screening Methods Available

Once the need for a screening programme has been established, the screening method must be considered. There are a wide range of imaging and non-imaging techniques which could be used for screening; these are considered and evaluated in the following sections.

1.2.1 Breast Self Examination

Breast Self Examination (BSE) is a regular (monthly) systematic examination of the breasts in order to detect lumps within the breast while they are still relatively small. The technique is cheap, the educational materials and teaching time being the major expense. Joensuu [31] found that as 10% of all cancers found in a screening programme had been discovered by the women themselves, the technique obviously has some value. The results achieved will inevitably depend on the proficiency with which the examination is performed [32]. It has not been shown to be particularly effective; Costanza [33] quotes no mortality benefit for breast self examination. When lesions are large enough to be palpable, they are generally invasive and well developed [32,34,35,36]. Baines [36] reviewed the literature and concluded that BSE is associated with a smaller tumour size at diagnosis but that it is uncertain whether or not this translates into a reduced mortality. This is not a suitable method for screening but is seen by Foster et al [34] as something which should be used as a complement to X-ray mammography.

1.2.2 Clinical Breast Examination

The breasts are examined by a trained practitioner, either a nurse or a doctor, on a regular basis. This can be done as a stand alone procedure or in conjunction with other screening techniques. This was used as one of the arms of the Edinburgh trial [37] but its value is still unproved. Essentially it has the same limitations as Breast Self Examination but without the advantage of very frequent examination. As a stand-alone screening test it has little value but can increase the sensitivity of a screening session when used in conjunction with, for example, X-ray mammography. In Japan, where the sensitivity was 61.1% and the specificity was 94.5% (n=8271), Joensuu [31] recommended that X-ray mammography should be used for screening. Winchester [35] recommends combining X-ray mammography and clinical examination.

1.2.3 X-ray mammography

This is the examination of the breast by X-ray imaging. The breast is an object which has intrinsically low contrast to X-rays and so highly specialised techniques such as the use of low X-ray tube voltages, a molybdenum target, K-edge filtration and compression of the breast are used. These will be discussed further later in section 2.4.1. This is currently the method of choice for screening, recommended by the World Health Organisation [38]. The disadvantages are that it is uncomfortable, there is a small but significant risk of carcinogenesis associated with the radiation dose received and the images produced are difficult to interpret, requiring specialised training of both radiographers and radiologists. The sensitivity in a screening situation has been found to be 93% and the specificity 95% (first round of screening in NE Thames).

1.2.4 Xero-mammography

This is very similar to X-ray mammography, however a Xerox selenium plate is used as the image receptor, and higher X-ray tube voltages are used than with conventional mammography. The main advantage is that the images show edge enhancement. The technique has fallen out of use due to the unacceptably high radiation dose involved [39].

1.2.5 Ultrasound

Ultrasonic examination of the breast has a resolution limit of around 3mm but has none of the risks of ionising radiation and is a relatively simple, if time consuming, examination. On its own, it has a low specificity and is therefore unsuitable for screening. However, it is a useful adjunct to X-ray mammography and is particularly good for distinguishing between cysts and solid lesions and to a lesser extent, between malignant and benign tumours. There is also a role in examining symptomatic women under 30 years of age, women with breast lumps in pregnancy, women with breast tenderness and women who do not wish to be exposed to ionising radiation. Guyer [40] reports using Doppler ultrasound to look at the blood flow and reports finding increased vascularity in most cancers. There is some concern about the biological effects produced by ultrasound [41,42] but current information indicates that only high intensity techniques such as pulsed Doppler are a cause for concern.

1.2.6 Thermography

Thermography is a technique where the temperature of various parts of the breast is measured from the black body radiation emitted. There are two assumptions made in the use of thermography for breast disease, firstly that the normal condition of the body is to have a symmetrical heat distribution and secondly that breast tumours tend to be heavily vascularised and show up as hot spots. Thermography is slow and difficult to perform due to the need to achieve a steady temperature in the room in which it is performed. The main disadvantage is that it is only able to look at the surface of the breast. There is a considerable range of temperature within a healthy population, and abnormality can be defined as results which lie more than 2 standard deviations from the mean of a normal population [43]. This is also complicated by the fact that there are many reasons for increased temperatures other than malignant breast disease and some types of malignant disease which do not produce a rise in temperature. Consequently the sensitivity and specificity are low (61% and 74% respectively [44]). The same study also indicated that the test was not useful as an indicator of the likelihood of developing breast cancer.

1.2.7 Transillumination

This technique involves shining light through the breast and looking for areas of increased attenuation. It has very little diagnostic value and has not been used for many years.

1.2.8 Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) is a technique which uses radio frequencies to examine the chemical composition of objects [45]. It has no known adverse effects and is highly sophisticated, and therefore very expensive. It is also slow to perform a scan taking around twenty minutes. Groups in Liverpool, Lewis-Jones [46] and Hickman [47] and Oxford, Westbrook [48] have successfully used MRI with Gadolinium DTPA enhancement for looking at recurrences of previously known breast cancer and differentiating between cancerous tissue and scar tissue in equivocal screening cases respectively. The Liverpool group do not advocate MRI as a screening technique and Westbrook states that it is unlikely that MRI will replace X-ray mammography for screening

purposes. From the Liverpool studies, the sensitivity is 100 % and the specificity is 94 %; this is in a population where the incidence of malignant disease is much higher than in the general population. Breast coils are being developed by the manufacturers of MRI equipment, who obviously predict increased use for breast work in the future. As new faster imaging sequences are developed, the use of MRI for breast cancer screening becomes a possibility.

Heywang-Köbrunrner in Leipzig [49] has tested contrast enhanced MRI and found a sensitivity of 99.5%, specificity of 28%, a PPV of 61% and a NPV of 98%, and states that although MRI could not be used as a primary screening method, it has the ability to improve diagnostic accuracy in a population which has been screened by mammography, and is particularly useful where a scar is present from previous surgery or biopsy, where the woman has a silicone implant or where the breast tissue is very dense.

1.2.9 Digital Mammography

This technology is very similar to conventional mammography, the main difference is the storage medium for the intermediate X-ray image; the X-rays exiting from the breast excite a storage phosphor, the data stored on the phosphor is subsequently digitised by a laser reader which stimulates the phosphor to release the stored energy [50,51,52]. The main limitation is due to pixel size which means that the smallest of calcifications cannot always be imaged. The trade-off is that low and medium spatial frequencies exhibit a better detected quantum efficiency. Parkin [53,54] reports that the image quality is as good as conventional mammography and superior for dense breasts. The fact that the image is stored digitally means that data transmission, archiving and retrieval of images are all simpler and less time consuming. There is the ability to manipulate the image to enhance the performance of the radiologist and there is room for some dose reduction, although in many cases, conventional mammography is limited by quantum mottle and thus a dose reduction would not be possible in these cases.

There have also been attempts to automate reading by computer analysis of digitised images but so far these have proved to be of little value [55]. With screening organised in its present form, this would also be prohibitively expensive as a reader would be needed on each mobile unit,

or alternatively several hundred storage phosphors would need to be available on a mobile unit and transported back to base to be read.

1.2.10 Computed Tomography

Although computed tomography (CT) is particularly good at imaging small variations in attenuation which would make it ideal for the breast, this technique gives a very large radiation dose and consequently could not be justified in terms of the risk benefit ratio. However, there is a potential role for CT during assessment where Silva [56] has reported its usefulness, when used with a contrast agent, in differentiating between fibroadenomas and cancers when conventional imaging is unable to do so.

1.2.11 Genetic Markers

Garret et al [57] have identified HRAS alleles which are associated with the presence of breast cancer; this link is stronger in black women than in white. Powles [58] and Lalle [59] have pursued the idea that breast cancer has a family link. When they studied high risk families, they found evidence of a BCRA1 germline mutation which, it is thought, predisposes women to breast cancer. Although there are currently no tests suitable for screening using genetic markers, it is probable that such tests will become available in the future. A genetic predisposition to breast cancer does not mean that a woman *will* develop a cancer, merely that if the woman is exposed to a carcinogenic agent that she is more likely to develop a cancer than other woman. Neither does this exclude the possibility that a woman who has no genetic predisposition will develop cancer; if enough carcinogenic exposure is received, a cancer will be generated.

1.2.12 Thermal Bra

A novel piece of work by Simpson and Griffiths [60,61] has highlighted the possibility of direct breast temperature measurements using temperature sensors within a "Chronobra". They found that high risk women (selected because they had already had surgery for breast cancer in one breast) did not show the temperature changes associated with the menstrual cycle which is exhibited by the normal risk control group. Although this technique would be of limited use for the mainly post-menopausal 50+ age group currently involved in the breast screening programme, it has no

side effects and might be useful for younger women who are not suitable candidates for X-ray mammography because of the radiation risk.

1.3 Survey of Current Screening Programmes

1.3.1 USA

Screening is performed in the private sector and there is no government policy to provide breast screening as a right. Consequently access to screening is restricted to the more affluent members of society. Annual clinical examination and mammography for women over 50 and annual clinical examination and biennial mammography for women 40-49 years of age are recommended following a baseline mammogram at the age of 40. A major problem seems to be achieving good compliance [33].

There are guidelines from the various professional bodies and screening clinics are strictly regulated if they wish to be accredited. The American College of Radiologists run an accreditation programme to ensure reasonable quality in mammography [62] and it is thought that this will become mandatory at some point in the future.

Breast self examination is recommended

1.3.2 Australia

In Australia, there is no national screening programme as such, but many states are implementing mammographic screening. In Western Australia [63], the programme has begun in the cities and is expected to have been implemented state-wide by 1995. The target group are in the age range 45-69, screening is done every two years and two views are taken. Major problems are logistical in nature because the state of Western Australia incorporates a very large area which is sparsely populated.

Breast self examination is recommended.

1.3.3 Iceland

The programme is country wide and implemented by the government. A baseline screening mammogram is done at age 35 then mammography is done every 2 years between the ages of 40 and 69 [14].

1.3.4 Ireland

Screening is still in the pilot stage with a two arm study in progress [64] but is likely to be implemented throughout the country in the near future in a form similar to that in the UK.

1.3.5 Holland

There is a well established screening programme based upon the research at Nijmegen and Utrecht. The screening is done by mammography every two years and women from the age of 50 and 70 are eligible. The whole country is covered and it is centrally co-ordinated [14,65].

1.3.6 Sweden

From the ages of 40 to 54, mammography is done every 18 months and from 55 to 74 mammography is done every two years. This is a shorter interval than the one used in the two-counties trial where 2 years and 33 months were the intervals for the younger and older age groups respectively. There were suspicions that the screening interval being too long accounted for the lack of benefit shown in the under 50 age group [14].

1.3.7 UK

It is expected that a fully operational screening programme will prevent approximately 25% of breast cancer deaths within the screened age group (50 to 64 years) which represents one third of the annual total of breast cancer deaths, giving an estimated 1250 lives saved per annum [66]. This assumes that seventy percent of the invited population will actually attend for screening, as the acceptance rate falls, so too will the number of cancers detected and the number of deaths prevented.

In 1986, a document to the Department of Health was published [2] which took these factors into account and combined with results from the HIP study [67,68], the Swedish two-counties trial [69,70,71] and the Dutch trial [72,73] and preliminary results from the Guildford and Edinburgh trial in the UK [74] and concluded that deaths from breast cancer in women aged

50-64 years who are offered screening by mammography can be reduced by one third or more.

Breast self examination is recommended (breast awareness)

1.3.8 Finland

Mammography is performed every 2 years after the age of 50, at the moment the schedule goes up to 63 years of age but there is to be no upper age limit when screening is fully implemented [14].

1.3.9 Canada

In Canada, the screening programme does not yet cover the whole country. Of twelve provinces, one screens women annually by mammography and clinical examination from the age of 40 onwards and five screen every two years by mammography with clinical examination between the ages of 50 and 69. Breast self examination is included in the guidelines [14].

The wide range of screening regimes reflects the lack of certainty as to which age group to target, and how often screening should be done. There is also a strong influence due to the way in which healthcare is funded, for example in the United States the patients pay, either directly or from an insurance policy, for all healthcare which they receive; a cynical view is that it is in the financial interests of the medical profession to encourage frequent screening from an early age. In contrast, in the UK nearly all healthcare is publicly funded and there are many conflicting claims on resources; the UK breast screening programme targets women of a smaller age range and screens only every three years positioning itself at the opposite end of the range of options available from current evidence.

1.4 Strategies for Implementing a Screening Programme

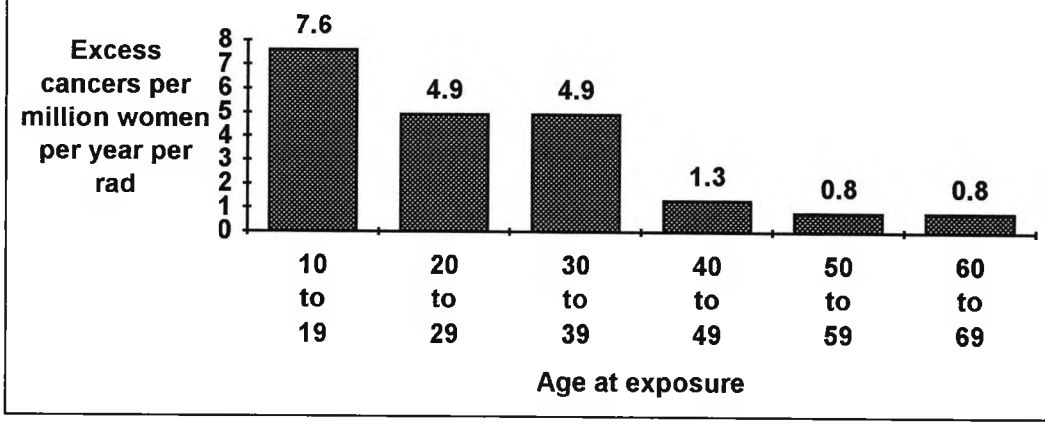
The word screening, when unqualified, implies that a test is done on the population as a whole. When the whole population does not suffer equally from the disease in question, and other factors, such as sensitivity to ionising radiation, come into play, the risk-benefit ratio changes for different groups within the population. In such circumstances, it is not only sensible but ethically necessary to examine the risk-benefit ratio and use it to select groups of people suitable for screening excluding those who are known to suffer rarely, if at all, from the disease in question, for example males who seldom suffer from breast cancer. It also makes sound economic sense to attempt to identify high risk groups and offer them greater resources than the general population.

1.4.1 Risk Factors

If there are risk factors associated with a disease, it is logical to try and target screening on the high risk group. The most important risk factors for breast cancer are being female and increasing age. These are obviously going to have the strongest bearing on selection of a target population [75] and will be discussed further in 1.4.2. Race and socio-economic status also have a bearing on the risk of developing and dying from breast cancer. Asian women in the USA generally have a lower risk of developing cancer although this increases in successive generations indicating that it is environmental rather than genetic factors which influence the risk [76], this does raise the possibility of some degree of cancer prevention by means of health education programmes if the key environmental features can be identified; Afro-American women have a particularly high risk of breast cancer mortality though this is not reflected in the incidence, this is thought to be due to poorer access to health care in this population group [76]. A close family history of breast disease is a risk factor for early occurrence of breast cancer, a gene has been tentatively identified as the carrier for this risk [77], it is however responsible for only a small proportion of breast cancers. Age at menarche, childbearing history, being overweight, moderate to high alcohol consumption (14 units per week, increases the risk of breast cancer by 1.7 [78]) and oral contraception are also risk factors, but they are relatively minor. Vessey [79] documents a slight increase in risk of breast cancer following hormone replacement therapy and occurring after a latent period of ten years.

Radiation exposure is a proven risk factor; data from the Japanese survivors of the atomic bomb, radiotherapy for post-partum mastitis and data from fluoroscopy used for TB screening show that the sensitivity to radiation decreases markedly with age. This is illustrated in figure 1.4.1. For any dose below a mean glandular dose of 1 Gy the risk is zero or statistically weak [80]. The dose received in X-ray mammography is now typically 2 mGy per view.

Figure 1.4.1 Sensitivity of the breast to ionising radiation with age
Generated from data in reference [80]



1.4.2 Target Population

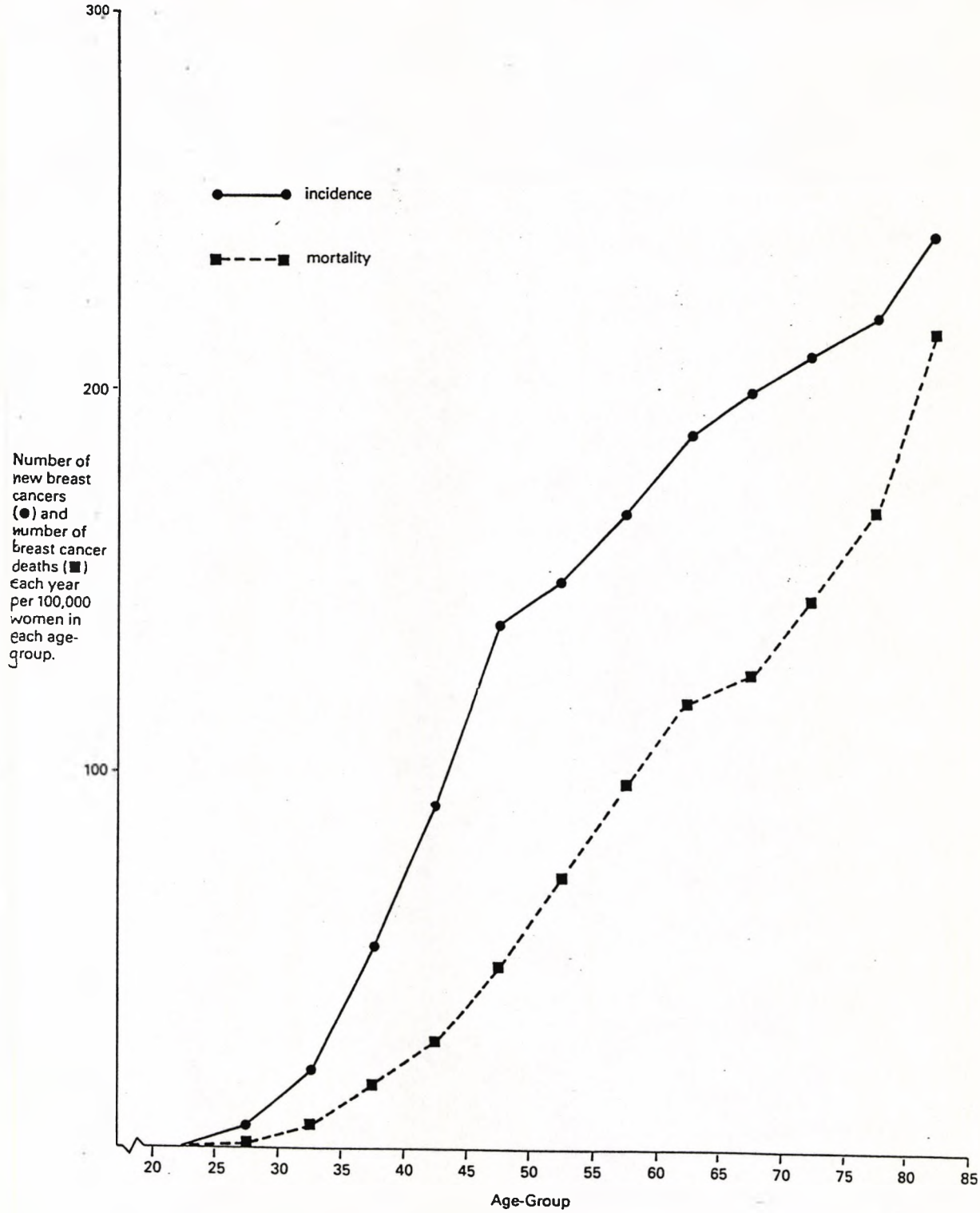
In order to most effectively target a group of women for breast screening it is necessary to know how the disease is distributed through society and which groups will actually benefit from screening. Knowledge of the risk factors which have been discussed enables a group to be selected for screening.

Of all the factors mentioned, the only one which is suitable for selecting women is age, all the others allow too great a proportion of cancers to slip through. Since the incidence of cancer increases with age, figure 1.4.2, it is obviously sensible to concentrate on older women where efforts will be more effective. However, although the incidence of cancer is usually higher in the older age groups, recent data from the USA shows a decrease in breast cancer incidence above the age of 75 years [75]. This has been ascribed to the effect of screening. This raises the question " should there be an upper age limit for breast screening ?". It is important to consider what is the mortality for the older age groups; if breast cancer in older women is slow growing then the option of doing nothing is a valid one and if there is to be no treatment following the discovery of a cancer, then there is no point in screening. One must also take into account co-morbidity of older women [81], if a woman is suffering from some life threatening illness and has a life expectancy of a year or two, and the lead time for breast cancer is of that order, again, screening is pointless, it simply adds worry to a woman in her last years. As life expectancy increases with advances in medical technology, there is a greater need to continue screening beyond the common 65 or 69 year age limit.

The increase in the incidence of cancer with age does not mean that younger women do not get breast cancer and there has been a heated debate about the merits of mammography for women under the age of 50. It has been stated (erroneously) that pre-menopausal breasts cannot be imaged well using mammography. This is not the case, however it is true that the risk of radiation damage is much greater in pre-menopausal breasts and that studies [82,83] show little or no benefit to women under 50. Heilbrunn [84] has reported a retrospective study of women in the 40-49 age group in the USA and found that approximately 70% of the cancers detected were stage 0 or 1 with a very favourable prognosis. The discrepancy with results from other studies has been attributed to the improvements in X-ray mammography since 1985.

Figure 1·4·2 Age-specific incidence rate (1982) and mortality rate (1985) for breast cancer, UK

Taken from reference 2



The genetic link is currently subject to a great deal of investigation [27,57,58,59,77,85,86] and the weight of evidence may eventually lead to screening by mammography being streamlined into a process which occurs for a group of women identified by a blood test as genetically at risk, alternatively screening could be augmented to include those younger women who are shown by genetic testing to be at risk. At the time when breast screening was set up in the UK this link was suspected from epidemiological data but was so weak that it did not justify inclusion in the breast screening programme.

1.4.3 Screening Interval

The screening interval will inevitably have a great effect on the number of cancers occurring in between screening rounds (interval cancers) and upon the risk of radiologically inducing a cancer; this needs to be determined on the basis of cancer growth rates.

Breast cancer is one of the slower growing human cancers, certainly in the clinical phase, but is extremely heterogeneous showing a wide variation not only in growth rate but also in metastatic pattern. Conant et al. [87] estimated doubling times for mucin containing carcinomas based on successive radiographs and found a range from 134 days to 636 days with a mean of 302 days. Henderson [88] postulates that the time gained by using X-ray mammography rather than waiting for the cancer to become palpable can be approximately calculated by assuming that :

a) mammography is able to detect cancers at about the 21st doubling

b) a cancer cannot be detected by touch until it is 10mm or greater in size, which is the time when a woman first becomes symptomatic; this corresponds to 10^9 cells present or 29.9 doublings

From these assumptions, table 1.4.3.1 has been constructed, the maximum lead time gained by mammography is in the last column, this is the difference in time for the two different detection thresholds. The choice of screening interval has two conflicting demands, the first is to screen as often as possible to gain the maximum time, if every woman was screened every day, all cancers would be detected just as they reach the beginning of the lead time, when they go through the 21st doubling, giving

a time gain equal to the maximum lead time for every woman. The other demand is that screening is done as seldom as possible because:

a) it costs less
b) there is less cancer inducing radiation dose accumulated.

Doubling Time (days)	Time to 21st doubling (Detectable by mammography)	Time to 1cm ~ 10^9 cells (Detectable by examination)	Maximum time gain (days)
10	210	299	89
20	420	598	178
30	630	897	267
40	840	1196	356
50	1050	1495	445
60	1260	1794	534
70	1470	2093	623
80	1680	2392	712
90	1890	2691	801
100	2100	2990	890
110	2310	3289	979
120	2520	3588	1068
130	2730	3887	1157
140	2940	4186	1246
150	3150	4485	1335
160	3360	4784	1424
170	3570	5083	1513
180	3780	5382	1602
190	3990	5680	1690
200	4200	5979	1779

Table 1.4.3.1 Maximum possible time gain benefit of screening as a function of doubling time

It is instructive to consider the consequences of selecting screening intervals which are shorter than, the same as and longer than the lead time. If the following assumptions are made:

a) new cancers are arising at a constant rate of N per day
and

b) all cancers grow at the same rate (which is obviously incorrect in general, but will be approximately true for a particular type of cancer and simplifies the analysis)

it is possible to calculate the benefit of various screening intervals in terms of the number of days gained by early diagnosis per cancer. This is effectively the *mean* lead time per cancer, the lead time in table 1.4.3.1 is the maximum lead time possible (and would only be equal to the value of the mean lead time if screening were done every day).

First, if the screening interval is less than the maximum lead time, the detection of all cancers will be advanced by some lead time due to the screening process. The ones to gain least are the ones detected one screening interval (SI) after the day when they would become screen detectable, the ones to gain the most are those which are detected immediately they are detectable by mammography after 21 doublings. The cancers which gain *least* have a benefit of LT-SI days so the total benefit for such cancers is N*(LT-SI) days.

The cancers which are detected one day prior to this in their development show a benefit of N*(LT-SI+1) days due to early diagnosis, the next group obtain a benefit of N*(LT-SI+2) days and so on. This series continues up to a maximum of N*LT days gained for cancers which are detected immediately they enter the "lead time" phase of their development.

The total days gained in one screening cycle is equal to the sum of this series

$$Days\ gained = \sum_{i=0}^{i=SI} N(LT - SI + i) \qquad \text{Equation 1.1}$$

$$Days\ gained = \frac{(SI+1)}{2}\left(2N(LT-SI) + N.SI\right) \qquad \text{Equation 1.2}$$

$$Days\ gained = \frac{N(SI+1)}{2}\left(2(LT-SI) + SI\right) = \frac{N(SI+1)}{2}(2LT - SI) \qquad \text{Equation 1.3}$$

and the mean time gained per cancer is this number divided by N*SI, the number of cancers occurring during one screening cycle, and is

$$Mean\ days\ gained = \frac{(SI+1)(2LT-SI)}{2SI}$$ Equation 1.4

The other regime is when the screening interval is greater than the maximum lead time. Some cancers will be detected by screening, but a proportion will have gone through their "lead time" phase and become clinically manifest before the next screen is performed, these are true interval cancers (as opposed to false negative cancers which were not detected by screening due to limitations of the technique, and are ignored in this analysis).

The cancers which show the least benefit are the ones which have gone beyond their lead time phase and are in the clinical phase at the time screening is performed. Their benefit due to screening is zero days. A benefit begins to be seen when the cancer is just pre clinical, when the benefit is 1 day times N cancers, this then becomes 2N days, 3N days up to a maximum of N*LT days when the cancers detected are just entering the lead time phase.

$$Days\ gained = \sum_{i=0}^{i=LT} iN$$ Equation 1.5

The summation of this series is given by

$$\begin{aligned} Days\ gained &= \frac{LT}{2}(2N+(LT-1)N) \\ &= \frac{LT.N}{2}(2+(LT-1)) \\ &= \frac{LT.N}{2}(1+LT) \end{aligned}$$ Equation 1.6

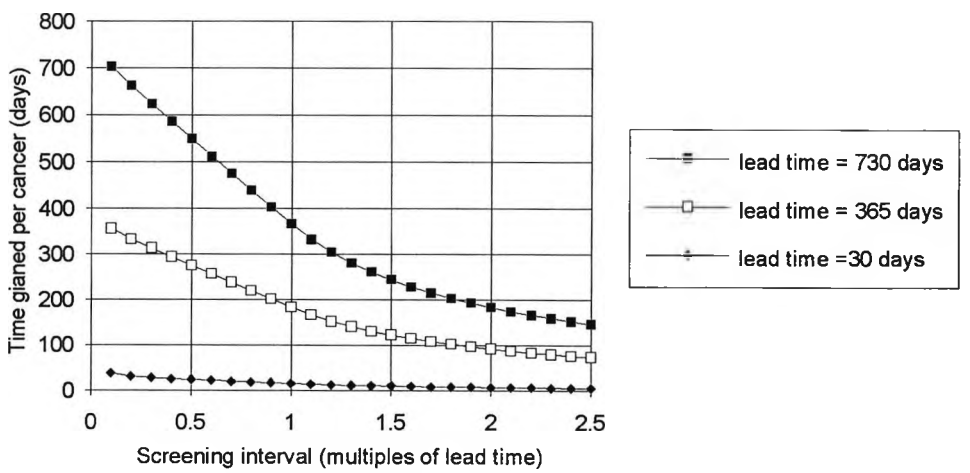
the number of cancers arising during a screening period is SI*N so the days gained per cancer is

$$Mean\ days\ gained = \frac{LT}{2SI}(1+LT)$$ Equation 1.7

The curves in figure 1.4.3.2 show days gained versus screening interval for a range of lead times. Matching the screening interval to the lead time gives a time gain per cancer of half the maximum (with daily screening) and no interval cancers. The detriment which is of most importance in determining the screening interval is the radiation dose; the dose increases linearly with the screening frequency (1/SI), the detriment

may have a threshold before any harm takes place or may be quadratic or linear-quadratic. Using a linear response model, the harm at low doses (where epidemiological data is sparse or non- existent) tends to be higher than any of the other models. i.e. a linear model gives the worst case scenario for low doses. The gradient of the line representing the radiation carcinogenesis detriment depends on the dose per examination and the risk of carcinogenesis per unit dose. In order to justify screening, the benefits of finding cancers early must outweigh the harm. A fuller discussion of the trade-off between risk and benefit is given in section 1.1.3

Figure 1.4.3.2
Mean time gained per cancer due to screening as a function of the screening interval



In a more realistic situation, the doubling rates vary greatly and the detectability of a cancer varies depending on the type. Some cancers can be clinically detected when smaller than 1 cm, others can never be clinically detected even when they are large, and some cancers are not detectable at all by mammography. Nevertheless, the calculations form a useful basis for making this type of decision.

An alternative approach is to base the chosen screening interval on experimental data from randomised trials choosing the interval which produces the largest reduction in breast cancer mortality. There are a number of confounding factors in any trial (for instance the attendance rate in the screened group) which make it quite difficult to separate out the effect due to screening interval. Although there are not yet any results from randomised studies, there are currently a number of trials running to address the effectiveness of different breast screening intervals [33].

1.4.4 Call and Recall

The ultimate aim in breast cancer screening is to reduce the mortality from breast cancer in the population as a whole. Women who accept screening tend to be more health conscious than those who do not and are more likely to be living a healthy lifestyle, this is called a self-selection bias and means that a screened group is likely to have a lower mortality than an unscreened group before the effect of the intervention is taken in account. The true benefit of screening must be measured over the whole population, and if the uptake of the offer of screening is low, the benefit is diluted. The corollary of the self-selection bias is that in the unscreened group, the breast cancer mortality is higher than average so increasing the uptake brings a disproportionate reward in the cancer detection rate. While it is important to respect the right of women to decline to be screened, it is equally important to encourage as many women as possible to attend, particularly those in socially disadvantaged groups who are in fact the group most likely to benefit.

In many cases, the main barrier to achieving good coverage of the population is lack of motivation on the part of the women rather than active resistance to screening. There have been many studies looking at techniques to ensure that as many women as possible are encouraged to attend [10,89,90]. In the UK a call and recall system based on medical registers is used to give every woman an appointment every three years. If she fails to attend or reschedule the appointment, she is then sent a second appointment.

There are obviously difficulties when women change the area in which they live but there is a "fail-safe" procedure in place which should ensure that no woman who is registered will have to go more than three years between screens. As a batch of screening is completed, the women registered in the area or GP practice which has just been completed have their screening record checked on the computer. Any women who have moved into the area within the last three years are then checked to see when their next screening episode is due to take place. In most cases, the previous screening examination will not have taken place exactly three years previously, so they would not normally be selected by the computer for screening; if there is an additional three year delay before those women are next invited, the time between the two successive screening episodes would be unacceptably long. Such women are invited for

screening at the end of the batch, this ensures that there will have been a gap shorter than three years since the previous screening examination.

1.5 Treatment

The predominant benefit of screening for breast cancer is the early detection of a cancer which makes the treatment easier and more effective. The type of treatment given can vary enormously due to a range of factors such as the preference of the woman, the facilities available in a particular area, and the surgeon's experience and preferences. Nominally identical treatments may vary somewhat particularly if an experienced breast surgeon is being compared to a general surgeon with no special interest in breast work.

It has been suggested by Spratt et al [91] and also by Badwe et al [92] that the day of the menstrual cycle on which surgery is performed influences the survival, survival is reduced if the surgery takes place between three and twelve days after the last menstrual period. The majority of women involved in breast screening are post menopausal, and consequently this is not a consideration, however, should the age at which screening commences be lowered, the timing of surgery would then become significant.

The use of tamoxifen, a drug which acts against oestrogen, in the treatment of breast cancer is currently undergoing clinical trials and is likely to become more common. Sagar and Lopez [93] offer the use of tamoxifen when axillary dissection, to remove any potential spread to the lymph nodes, is not performed as part of surgery. Dookeran et al [94] suggest that tamoxifen is most appropriate for grade I or II tumours.

Chemotherapy has been considered to be a last resort treatment for metastatic disease, but now, its use in combination with other forms of treatment is being evaluated. Rayter and Phipps [95] have successfully used chemotherapy to decrease the size of tumours prior to surgery, and in some cases to combine chemotherapy and radiotherapy and avoid the need for surgery altogether.

Choice of treatment cannot be considered in isolation from the staging of the cancer, the most appropriate treatment will vary depending on the type and staging of the disease. Both Rubens [9] and Amalvic et al [96] have found that a lumpectomy and adjuvant systemic therapy used for early breast cancer, i.e. localised to the breast with no signs of locally advanced disease, is as good (in

terms of mortality) as radical mastectomy. This would not be the case where the disease had spread to the axillary nodes.

In the UK, the British Association of Surgical Oncology (BASO) in conjunction with the Royal College of Surgeons has produced guidelines for use within the NHS breast screening programme [97] which recommend

- that the majority of cancers (over 60%) should have biopsy or fine needle aspiration to give a histological or cytological diagnosis of cancer pre-operatively to minimise unnecessary surgery
- that there should be education of surgeons within the breast screening programme to make them aware of non-surgical treatment options for certain types of lesions
- that breast conservation should be used wherever appropriate with a target of 50% of cancers less than 15mm in diameter to be treated using conservation methods
- that nodal status should be ascertained in order to make appropriate decisions on the need for adjuvant therapy
- that DCIS patients should be entered into clinical trials and that for such patients, radiotherapy is contra-indicated,

Chapter 2

The Role of Quality Procedures

2.1 Description of the Breast Screening Service

The production of a list of names and addresses of women eligible for screening begins with the Family Health Services Authority (FHSA), the organisation which maintains records of all people registered with General Practitioners. From the register, women between the ages of 50 and 64 are selected. Inevitably there will be some errors on the computer records due both to incorrect entries and, particularly in inner city areas where people tend to move home frequently, due to information which is out of date.

The FHSA computer generates 'prior notification lists' (PNL) of the selected women which are sent to the relevant general practitioner (GP) for correction. These corrections should include the removal of women from the list who have died or who have had a bilateral mastectomy, the correction of typographical errors and notification to the FHSA of women who have moved away with the new addresses if they are known. There may also be women who should be on the list whose names have been omitted, these names should be added to the list but in fact are unlikely to be unless the GP practice holds its records on a computer database. The PNL's are then returned to the FHSA where the appropriate amendments are made.

The FHSA sends the revised lists electronically to the screening office computer. Individual appointments are then sent out to the women in batches which are usually chosen geographically to match the location of the mobile screening units. Usually the mobile units travel to locations near where the women live and the films are processed at the end of each day at the assessment centre, a few units (only one in North East Thames) have a processor on board the mobile.

If a woman fails to attend for screening she is either sent a second appointment or sent a letter inviting her to make a further appointment depending on the operational policy of that screening unit. If this elicits no response, the screening episode (the record of the woman's progress through the current round of screening) is closed and no further action takes place until three years later when she will routinely be called for screening again.

Some women cancel their appointment because they do not wish to have mammography. Their screening episode is closed and they are invited again in three years time unless they have written to say that they do not wish to have mammography ever, in which case they are removed from the list.

The letters from a proportion of women who have moved and who have not informed the FHSA are returned to the screening office by the post office. This information is passed on to the FHSA and the GP concerned so that they can update their records and take any appropriate action.

There is also a group of women who change their appointment because it is not convenient. A new appointment is made on the computer and the screening process continues in the normal way.

Women who do attend for screening can be classified in four ways:

A Normal
B Suspicious
C Lesion Present
D Inadequate film

The women in group A are sent a letter informing them of their result and the episode is closed for routine recall in three years time.

Women in group D are called back for further films and then proceed in the same way as the original screen would have done if the films had been of adequate quality for diagnosis.

Women in group B are usually called for further films to be taken, second stage screening. This involves further more sophisticated views of the breast to enable a more accurate diagnosis to be made (the assessment process). This will often include the use of ultrasound and physical clinical examination. Because specialised equipment is required, this usually takes place at the assessment centre which is often housed in the same building as the breast screening administration. If the suspected abnormality is shown to be present at assessment and is palpable, the woman then goes for localisation of the lesion which may also include fine needle aspiration of cells and will certainly involve the insertion of some kind of marker into the breast enabling excision biopsy or surgery to take place afterwards. If the abnormality is not palpable,

further views will be required and localisation of the lesion is done under ultrasound or X-ray guidance.

If the suspected abnormality disappears when different views are used, it can be assumed to have been an artefact and the woman can then be told that she does not have breast disease and can be put back on to routine recall.

Occasionally the radiologist will be unsure even after further views whether there is any abnormality. In order to prevent unnecessary invasive procedures, the woman may be put on early recall (typically six months to one year later). If there is disease present, it will have changed in appearance during that time enabling a positive diagnosis to be made.

Women in group C who have a clearly identifiable lesion present do not need to have assessment, they can go straight on to localisation during which cells will be taken and the nature of the lesion diagnosed as benign or malignant. A significant number of these will be cystic in nature and can be treated by aspiration. Where stereotactic fine needle aspiration is not available wire markers will be placed in the breast to enable biopsies to be taken.

Although this is the most usual pattern of diagnosis it is not completely fixed, the woman's wishes will play a part in deciding how to proceed, as will as unusual clinical indicators.

The number of women following any given path depends not only on the incidence of the disease in the population but also on the technical skill of the screening staff, the experience of the radiologists, the performance of the equipment and the willingness of the women to co-operate with the programme.

Table 2.1 shows the proportions of women who have gone through each stage of screening for three centres, A an inner city service, B a suburban service and C a rural service in the first round of screening.

Points to note:

1 The uptake of the service (number attending for screening/ number invited) is exceptionally low at centre A and yet the impact on the screened population (the percentage of cancers detected) is greater than that at the other two centres. This would seem to indicate that there is:

a) over diagnosis at centre A,

b) under diagnosis at centres B and C or

c) centre A is exhibiting a selection bias (those most at risk are those most likely to attend for screening).

2 Centre A biopsies nearly twice as many women out of those screened as do the other two centres, and yet detects only 25% extra cancers. It would appear that the additional biopsies produce diminishing returns in terms of cancers detected.

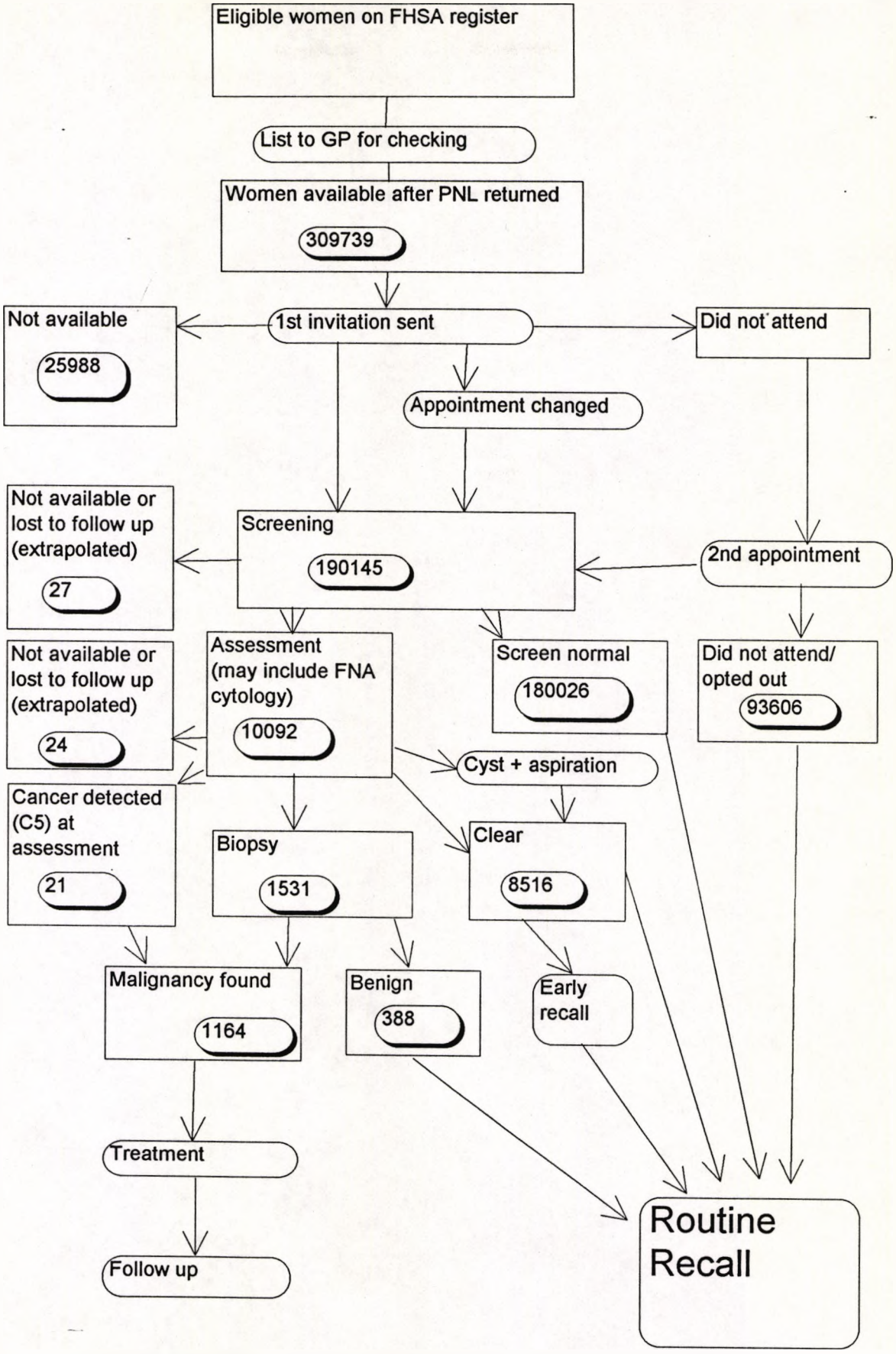
	A - inner city	B - suburban	C - rural
Number invited	31154	38570	45540
Number screened (% of those invited)	13029 (42%)	28476 (74%)	34350 (75%)
Number assessed (% of those screened)	912 (7.00%)	1541 (5.41%)	1551 (4.52%)
Number biopsied a (% of those screened) b(% of those assessed)	186 a(1.43%) b(20.4%)	195 a(0.68%) b(12.7%)	260 a(0.76%) b(16.8%)
Cancers detected a(% of those screened) b(% of those biopsied)	109 a(0.84%) b(58.6%)	161 a(0.57%) b(82.6%)	197 a(0.57%) b(75.8%)
Benign to malignant biopsy ratio	0.71:1	0.21:1	0.32:1

Table 2.1 The variation in screening performance in different environments

Figure 2.1 is a flowchart showing the numbers at each stage of screening. The figures used are the total for all of the units in North East Thames for the first round of screening. Several data items, the number of cancers classified as C5 during assessment, the number of interval cancers and the numbers of women not available at the assessment and biopsy stages, are not available from the statistical returns and have been estimated by extrapolating hand held data covering a period of approximately twelve months.

Figure 2.1

North East Thames RHA Breast Screening Programme



2.2 Evaluating the Decisions - a statistical description of the screening process

At every stage of screening, decisions are made as to whether the client is healthy or requires further investigation. The decisions can be classified in four ways:

1 There is disease present and it is correctly diagnosed (true positives, the indicator of which is taken to be the number of histologically positive cases)

2 There is disease present and it is diagnosed as clear (false negatives, the indicator of which is the number of cancers which were already present at screening). This information is not available from the computer system, in order to provide an estimate, the number used was the number of interval cancers divided by two, which is likely to be an over-estimate rather than an under-estimate.

3 There is no disease present and the diagnosis is negative (true negatives, this is approximately the number of women who are passed to routine recall, the number of true negatives is the number passed to routine recall minus the estimated number of false negatives).

4 There is no disease present and the diagnosis is positive (false positives, this is indicated by the number of women who go on for further investigation but are found to be clear of disease at a subsequent stage of screening)

One has to assume that the diagnosis made from biopsy is 100% accurate in order to be able to say which of the positive diagnoses are true or false. The following indices are used to evaluate the quality of the decision making [98].

$$\text{Sensitivity} = \frac{\text{Number of true positives}}{\text{Number of true positives} + \text{false negatives}}$$

$$\text{Specificity} = \frac{\text{Number of true negatives}}{\text{Number of true negatives} + \text{false positives}}$$

$$\text{Accuracy} = \frac{\text{Number of true positives} + \text{true negatives}}{\text{Total number of results}}$$

$$\text{Positive predictive value} = \frac{\text{Number of true positives}}{\text{Number of true positives} + \text{false positives}}$$

$$\text{Negative predictive value} = \frac{\text{Number of true negatives}}{\text{Number of true negatives} + \text{false negatives}}$$

All of these depend upon the decision threshold of the decision maker, the radiologist. If the decision maker is told that it is vitally important not to miss a positive result, the sensitivity will go up as the number of true positives increases but the specificity will go down as the number of false positives increases. The accuracy and positive and negative predictive values may go up or down depending on the initial decision threshold.

Looking at the data available for the services within North East Thames region (the raw data is available in Appendix B), the parameters at each stage of the screening process are as follows:

	Sensitivity	Specificity	Accuracy	PPV	NPV
Screening	0.925	0.953	0.953	0.115	0.999
Assessment	1.000	0.956	0.962	0.75	1.000
Biopsy	1.000	1.000	1.000	1.000	1.000

Table 2.2.1 Screening statistics for NE Thames region, Prevalent (1st) round

	Sensitivity	Specificity	Accuracy	PPV	NPV
Screening	0.875	0.968	0.968	0.111	0.999
Assessment	1.000	0.974	0.977	0.828	1.000
Biopsy	1.000	1.000	1.000	1.000	1.000

Table 2.2.2 Screening statistics for NE Thames region, part of Incident (2nd) round

There have been periods of time when an X-ray unit has been functioning sub-optimally, between a problem being identified and remedied. For one such X-ray unit, the screening statistics have been extracted for six months prior to a new tube being installed, a period when the resolution of the system was just below the requirement of 8.9 line pairs per mm set out in the Pritchard report [3], and a six month period after the tube was replaced when the resolution of the system was well above the requirement. There were other changes which accompanied the installation of the new tube, the output of the tube increased and simultaneously the AEC setting produced a lighter optical density when 4cm of perspex was exposed. When a test object was used with the new system, the image quality appeared to have increased. The main weakness of the data lies in the small numbers of women involved in a six month period, giving results of a very low statistical significance. Another weakness is that information on interval cancers is not available for short periods of time. Using a null hypothesis that the two sets of results are the same, the chi-squared test was applied to the data and gave a value of 0.725 $p>0.1$ for screening and 0.700 $p>0.1$ for assessment; this implies that the null hypothesis is true; the calculation is shown in appendix E. There are two possible conclusions which may be drawn from these results: the first is that the resolution of the system does not have an influence on the cancer detection rate, hence the non-significant result; the second is that the drop in optical density (which indicates a reduced dose and thus a worse signal to noise ratio if the system is quantum limited) cancels out the improvement in image quality due to better resolution.

Period	1st June 1992 to 16th December 1992	22nd December 1992 to 30th June 1993
Number screened	3681	4442
Number assessed	171	189
Number biopsied	22	15
Number of cancers	21	17*
Cancer detection rate	0.57%	0.38%
Sensitivity	100%	100%
Specificity	95.9%	96.1%
PPV	12.3%	8.9%
NPV	100%	100%

Table 2.2.3 Comparison of one unit with poor resolution and good resolution before and after change of X-ray tube.

* some cancers can be definitively diagnosed as C5 during assessment if fine needle aspiration of cells takes place.

2.3 Sensitivity analysis of changes in image quality

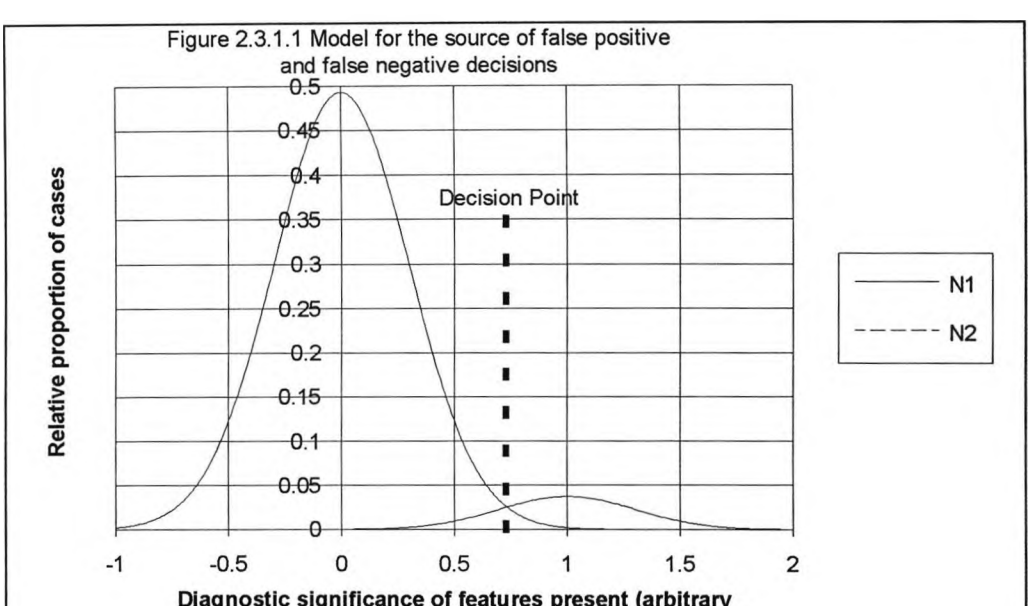
It is hypothesised that a decrease in image quality will have a detrimental effect on the final outcome of screening. If changes are to be made, or allowed to happen, to the breast screening system, then an evaluation of the likely effect of those changes is necessary. To do this, the screening system has been studied and modelled on a computer so that various parameters can be independently adjusted and the outcome of those adjustments observed. The cancer detection rate and the false positive rate at the initial screening stage have been identified as the main indicators of benefit and harm respectively, the analysis also looks at the number of cancers in the unscreened population and the false negative results produced by the programme as indicators of "lack of benefit" rather than actual harm.

2.3.1 Mathematical model for changes in image quality

When doing screening for breast cancer, one attempts primarily to discover whether or not there is a cancer present in the breast. The stage, size and development of the cancer is only of interest when choosing the way in which the disease is to be treated. The screening process is designed to divide the screened women into two groups: those with a cancer, and those without. It is inevitable that because of all the different types, stages and sizes of cancers that the number of diagnostic indicators will vary from one case to the next, and that some cancers will indeed have *no* diagnostic indicators and will be occult, whereas others may have so many diagnostic indicators that they are hard to miss. This model makes the assumption that the function of the number of women with cancer against the number of diagnostic indicators follows a normal distribution.

Likewise, the women who do not have a breast cancer present will have a range of diagnostic indicators, some of which may be due to benign breast disease, others of which may arise from the normal structures of the breast. It has similarly been assumed that these women form a normal distribution but that this distribution will be centred on a lower level of diagnostic features.

This is illustrated in figure 2.3.1.1 where the large curve centred on zero represents the women who do not have cancer and the smaller curve centred on one represents those women who do have cancer.



On average, women without cancer exhibit fewer diagnostic features on their mammograms than women with cancer, it is this fact which enables a diagnosis to be made at all. From figure 2.3.1.1, it can be seen that there is a considerable area of overlap - it is this overlap which gives rise to false positive results and false negative results. The job of the radiologist is to make a decision, based solely on radiographic features, as to whether each woman has cancer or not. Mathematically, this can be represented by a vertical line placed at the point along the x-axis (significance of diagnostic features present); this line represents the "decision point". Women whose mammograms display *more* diagnostic features than the number represented by the decision point are classified as *positive*, and women whose mammograms display *less* diagnostic features are classified as *negative*. The position of the decision point is entirely in the control of the radiologist who reads the film, and will be strongly influenced, consciously or subconsciously, by the perception of the radiologist of the role of screening. From the procedures adopted by many units in the UK such as double viewing where the films are evaluated independently by two radiologists, and two-view mammography which are not part of the original Forrest scheme [2], it would appear that radiologists perceive their job to be that of making sure that *all* breast cancers are detected by screening and none are missed - an impossible task. The danger of such an approach is that many more women will be called to assessment than necessary. The consequences of this are financial (to the screening service), psychological (to the woman) and the

increased risk of inducing cancers because of the increased radiation burden.

To counter this danger, the breast screening program produced targets in the Pritchard report [3], one of which was that the proportion of women going on to assessment should be less than 10% of those screened. This attempts to move the decision point further to the right.

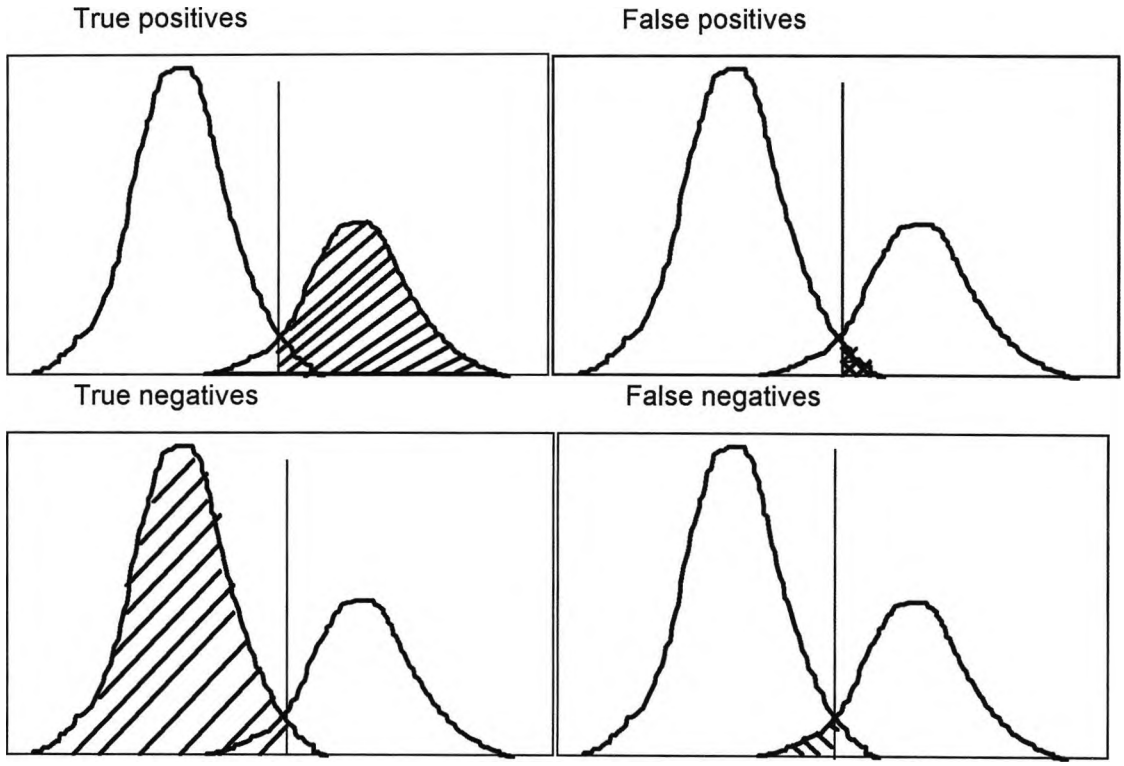
The relative height of each curve is determined by the incidence of cancer in the screened population, the area under each curve represents the numbers of women screened who do and do not have cancer respectively. The features which are germane to studying the role of physical quality assurance are the separation of the mean values of each distribution, and the spread of each distribution. In combination, these determine the degree of overlap, and consequently, the number of false positive and false negative decisions made, the false positive, false negative, true positive and the true negative decisions are shown in figure 2.3.1.2.

The further apart, and the narrower, the two distributions are, the easier it is to make a correct decision because there is less overlap of the two groups.

The x-axis represents the diagnostic significance i.e. the number and type of features seen. The variation in diagnostic significance for either group of women is due to two sources: firstly, the number of features which are actually there, and secondly, the number of features which the system (including staff and equipment) is able to detect. A perfect imaging system would produce no artefacts and be able to detect every feature which was present, and the spread of results would be due solely to the variations from woman to woman.

From this argument, it follows that the separation of the two curves is due to the distribution of diagnostic features within the screened population and is not influenced by the quality of the imaging process. It also follows that there will always be some false positive and false negative results, even when the imaging system is perfect, and that the numbers of false positive and false negative results will increase due to imperfections in the imaging system which widen the distributions.

Figure 2.3.1.2 How the number of true positive, false positive, true negative and false negative results are defined from areas under the curves



The model to represent the numbers of true positive, false positive, true negative and false negative results therefore consists of two symmetrical bell shaped curves at a fixed separation. It would be unreasonable to suppose that changes in image quality affect women with cancer differently from those without cancer therefore the spread of the curves σ_1 and σ_2 will be made to vary in the same manner for both cases and the heights of the curves will be chosen to ensure that the area under each one is proportional to the number of women with cancer and without cancer respectively. This assumes that the distribution of diagnostic features for either group of women forms a normal distribution.

The equation for a normal distribution is:

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad \text{Equation 2.1}$$

μ is the position of the centre;

σ is the standard deviation of the distribution.

This, of course, can be multiplied by a scaling factor so that the area under each curve represents the number of women with and without cancer respectively.

The cancer-free distribution, y_1, will be multiplied by a scaling factor of 0.993, and the cancer-present distribution, y_2, will be multiplied by a scaling factor of 0.007 to take account of the a priori probability of having cancer. The area under the curves will then be proportional to the number of women with cancer for y_2 and the number of women without cancer for y_1.

The values of μ_1 and μ_2 are arbitrary. It is the distance between them, $\mu_2 - \mu_1$ which is the factor determining the overlap of the curves. For the sake of simplicity, let $\mu_1 = 0$. This gives the equations:

$$y_1 = \frac{(1 - \text{incidence})}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}$$ Equation 2.2

$$y_2 = \frac{\text{incidence}}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma}\right)^2}$$ Equation 2.3

So the model has 2 variables, σ representing the inverse of the information content (or the *mis-information*) in the image, and x_d representing the decision point. The value of σ is determined by a combination of factors; contrast (a function of the film γ and the voltage across the X-ray tube), mean optical density (a function of the film speed and the mAs used for the exposure), the underlying structures seen (which depends upon what is actually present in the woman's breast, the number of views taken and the radiographic positioning) and the perception of the images (which depends upon the radiologist and the viewing conditions). μ_2 is a constant representing the mean differences in the underlying structures between women with cancer and those without cancer; it is initially unknown but can be determined from statistical data.

The first stage is to find values of μ_2, σ and x_d which give the proportions of true positive, false positive, true negative and false negative results which were experienced in the NE Thames screening programme. This was done using a spreadsheet to evaluate the functions for a range of values and using lookup tables to evaluate the integrals for the area under each curve.

The point of intersection of the two curves occurs when $y_1=y_2$, and can be found from the equation:

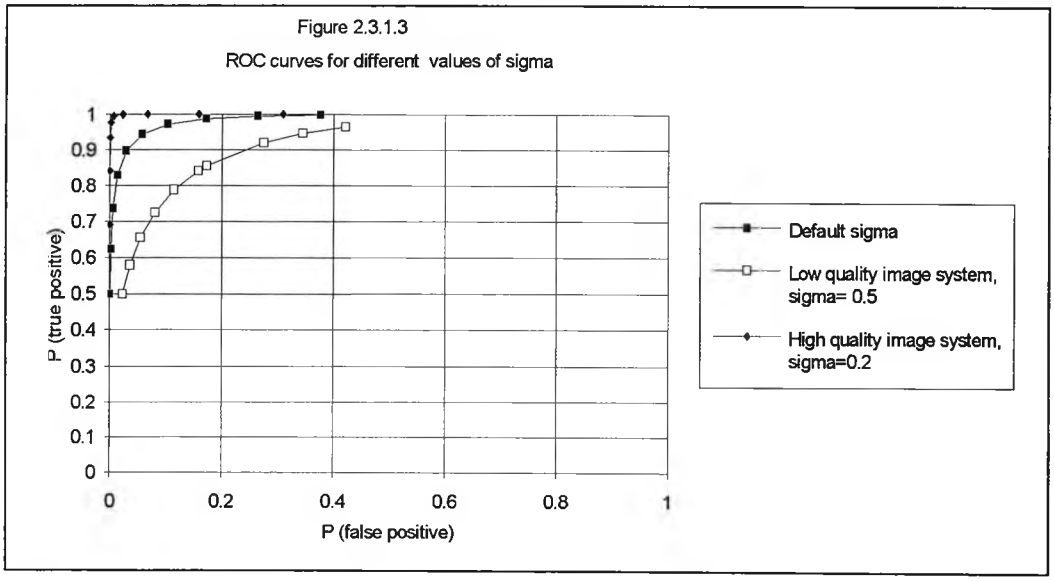
$$x = \frac{\mu_1 + \mu_2}{2} - \frac{\sigma}{(\mu_1 - \mu_2)} \ln\left(\frac{i}{i+1}\right) \qquad \text{Equation 2.4}$$

Where i is the incidence of cancer in the general population.

Initially, the assumption was made that the decision point and the point of intersection of the two curves coincide. For convenience, μ_1 was set to zero (it is the spacing between the distributions which influences the number of erroneous decisions, the absolute position does not matter). With this model, the value of μ_2 can be varied to change the spacing of the distributions and σ can be varied to change the information content of the images. A range of values of μ_2 and σ were tried and it was found that if they were varied in proportion with each other, i.e. when μ was doubled, σ was also doubled, the proportions of each type of decision, true positive, false positive, true negative and false negative remained the same. For the purposes of generating a model, μ_2 was set to a fixed value of 1.0 and σ was altered in order to change the degree of overlap of the two distributions.

Figure 2.3.1.3

ROC curves for different values of sigma



The effect of changing the value of σ can be seen in figure 2.3.1.3, each ROC curve has been generated for a particular value of sigma using the computer model, individual points on each curve correspond to different decision points. The default curve (in the centre) corresponds to the first

round screening outcomes for the whole of NE Thames, the other two values have been chosen arbitrarily for illustrative purposes, which might correspond to using two views for the examination [99] or a change in optical density [100]. The ROC curve is a standard method of describing the performance of an imaging system [136] the closer the curve is to the top left hand corner of the graph, the better is the imaging system. The term imaging system is used in its widest sense and includes the performance of the radiologist.

2.3.2 The results from the model

The numbers of true positive, false positive, true negative and false negative cases were taken from the screening statistics for North East Thames Regional Health Authority (NETRHA) and the variables σ and x (the decision point), were allowed to vary until the best least squares fit was found for the calculated results against the statistical data. It was found that the best fit occurred for the values $\sigma = 0.314$ and $x = 0.524$ representing screening, $\sigma = 0.358$ and $x = 0.609$ representing assessment and $\sigma = 0.100$ and $x = 0.500$ representing biopsy. Biopsy has been taken as the gold standard so a value of 0.1 for σ, the level of mis-information is at its minimum possible level.

The Bayesian likelihood matrix for screening can be calculated using the relationships [101]

$\pi_s L_{S/s} = G_{S/s}$ Equation 2.5

$\pi_n L_{S/n} = G_{S/n}$ Equation 2.6

$\pi_s L_{N/s} = G_{S/s}$ Equation 2.7

$\pi_n L_{N/n} = G_{S/n}$ Equation 2.8

Where the L values give the likelihood matrix, a set of figures describing screening, these values interact with the cancer incidence to generate the G values, or outcomes, which are the computer model generated equivalent of the screening statistics, π_s is the probability of cancer being present, (in non-mathematical terms, the cancer incidence, which was evaluated from the statistical data) and π_n is the probability of no cancer being present. π_s and π_n are not independent, there are only two possibilities, cancer is present or it is not, the probabilities must therefore add up to one.

$\pi_s + \pi_n = 1$ Equation 2.9

In a similar manner, the model was used to generated the outcomes and thence the likelihoods for the assessment stage of the process and the biopsy stage of the process. The biopsy stage was taken as the "gold standard" and therefore had a likelihood matrix of :

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

i.e. there are no false positives and no false negatives.

The entire process divides into three stages of decision making, screening, assessment and biopsy, represented by the Bayesian likelihood matrices which operate on the results of the previous stage. The end result is a spreadsheet on which a single parameter which influences the final outcome, can be altered, giving a final result indicating, for example, the numbers of cancers detected by breast screening. A full listing of the spreadsheet model is available in appendix A.

With this model, the following six parameters were altered independently:

1 cancer incidence,
2 the attendance rate,
3 image quality at screening
4 image quality at assessment
5 the decision point at screening
6 the decision point at assessment

This allowed predictions of how the number of cancers which are detected, missed, or incorrectly diagnosed are affected by changing these parameters.

Figures 2.3.2.1.1 to 2.3.2.6.4 show how the number of cancers changes as a result of changing each of the above parameters. It can be seen that the image quality and the decision point at screening have a very marked effect on the number of false positive and false negative decisions.

As might be expected, changing the cancer incidence produces linear changes in all of the measured outcomes as does the change in the rate of non-attendance (DNA rate). The reason for this is quite simple, the screening programme can only detect cancers which are there and in women who allow themselves to be examined.

When changing the image quality a small value of sigma represents a high image quality. At screening, the most dramatic change is seen in the number of false positives which rises sharply as the image quality is reduced (figure 2.3.2.3.1); simultaneously the cancer detection rate is reduced. Changing image quality at assessment alone has a similar result but on a much smaller scale. This is because the assessment process is performed on a much smaller fraction of the population. As one would expect, the number of cancers in the unscreened population (the fourth graph of each set) is unaffected by the image quality at any stage.

Changing the decision point at either screening or assessment illustrates the dilemma which is characteristic of any screening programme. Without an improvement in the ability of the test to discriminate between the two groups, any increase in cancer detection rate is paid for in an increased false positive rate.

Figure 2.3.2.1.1

Effect of changing cancer incidence on false positives at screening

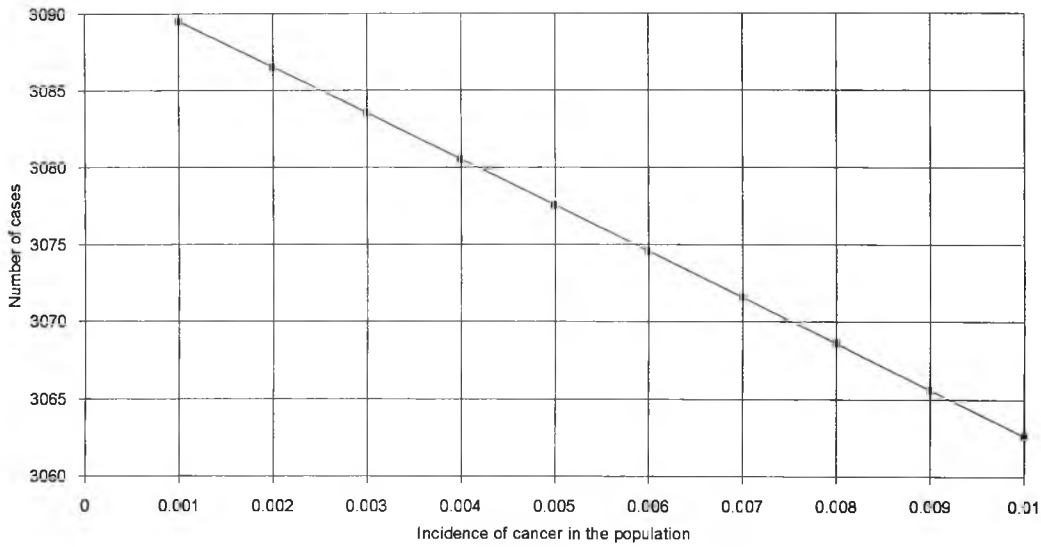


Figure 2.3.2.1.2

Effect of changing cancer incidence on number of cancers detected

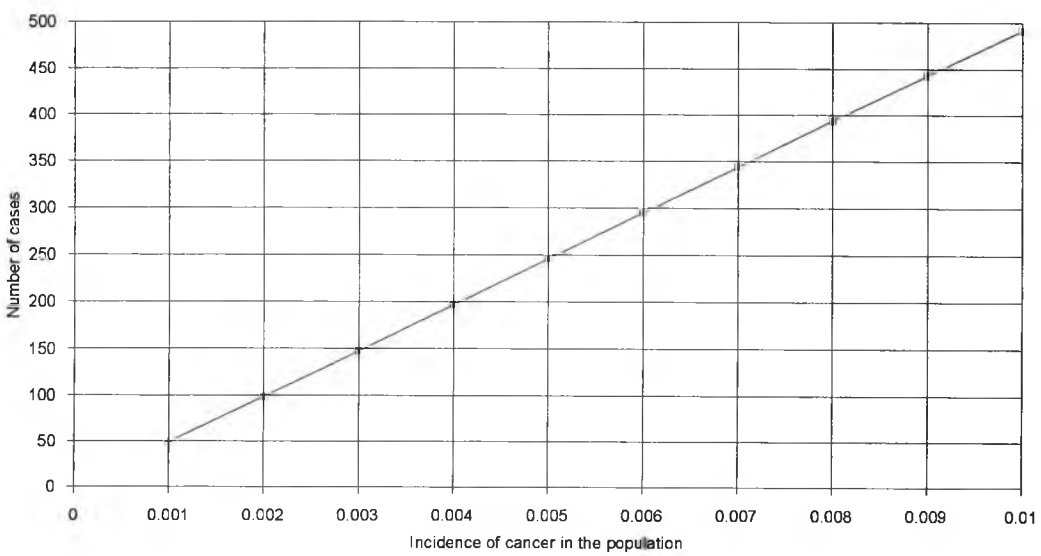


Figure 2.3.2.1.3

Effect of changing cancer incidence on number of false negatives

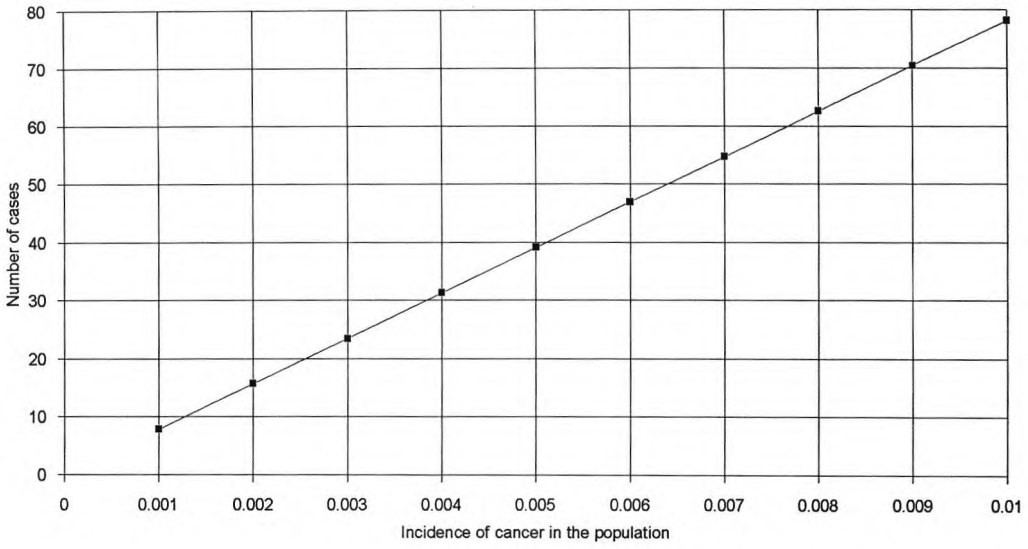


Figure 2.3.2.1.4

Effect of changing cancer incidence on number of cancers in the unscreened population

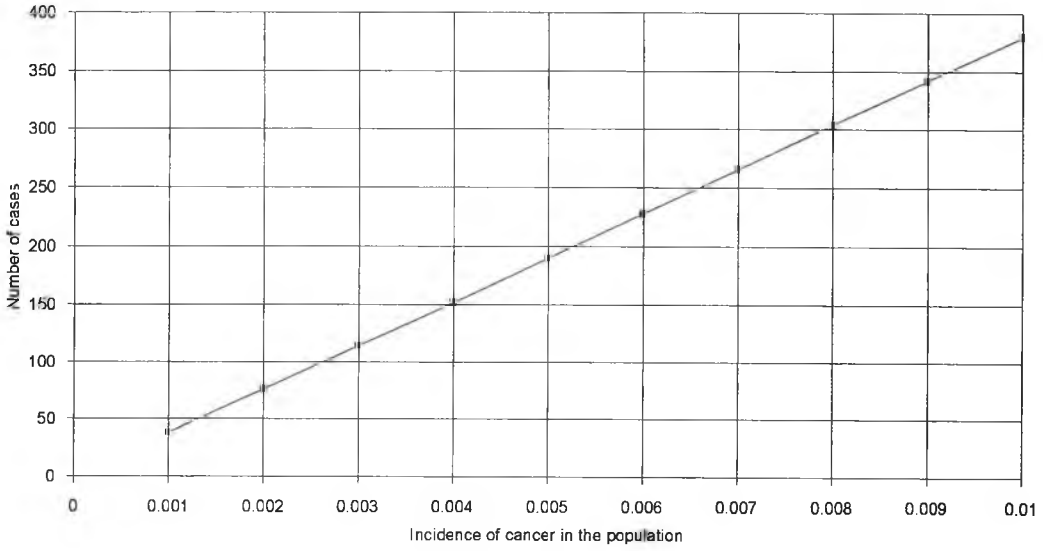


Figure 2.3.2.2.1

Effect of changing DNA rates on false positives at screening

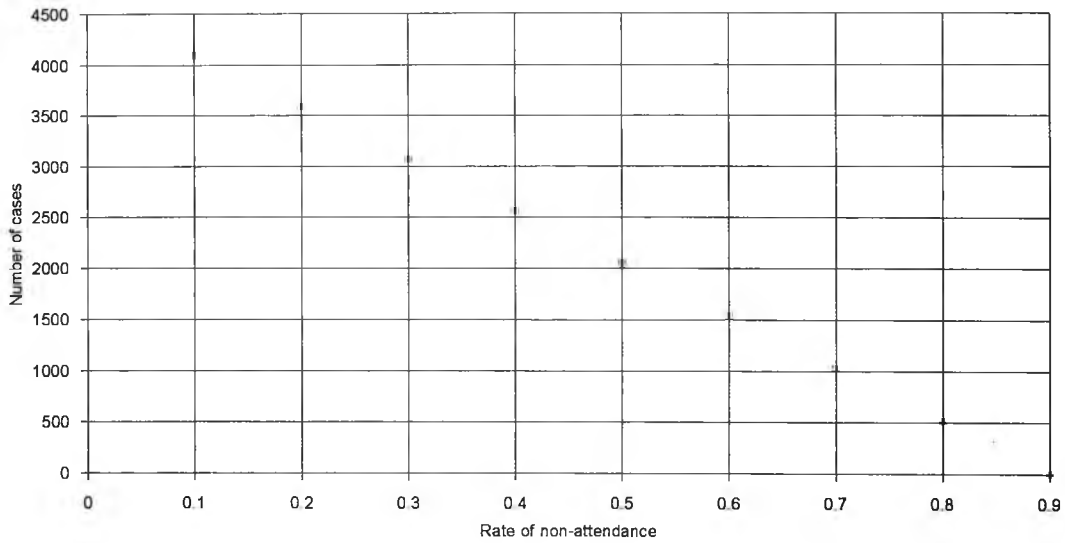


Figure 2.3.2.2.2

Effect of changing DNA rates on number of cancers detected

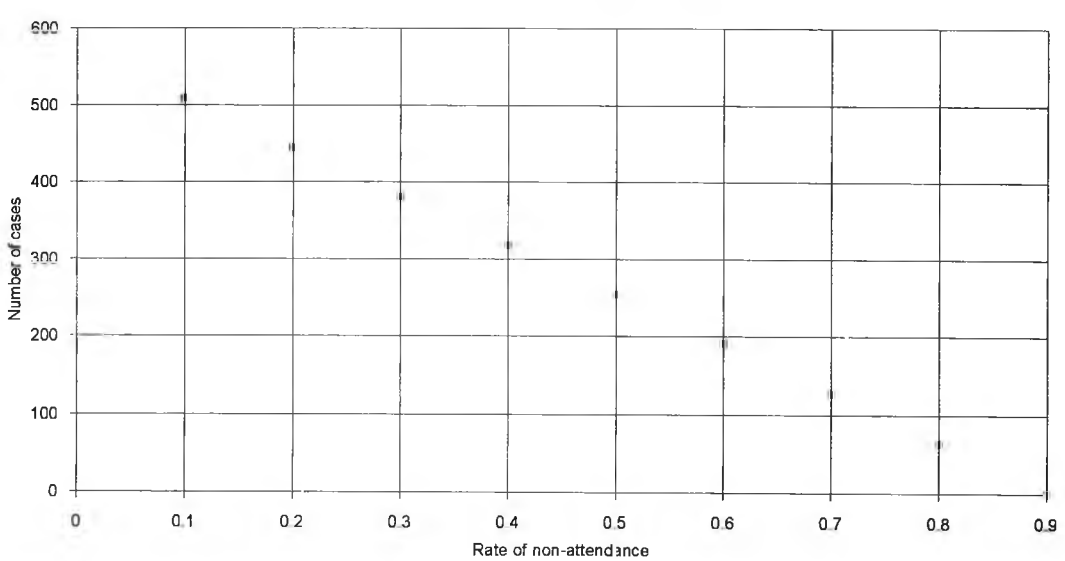


Figure 2.3.2.2.3

Effect of changing DNA rates on number of false negatives

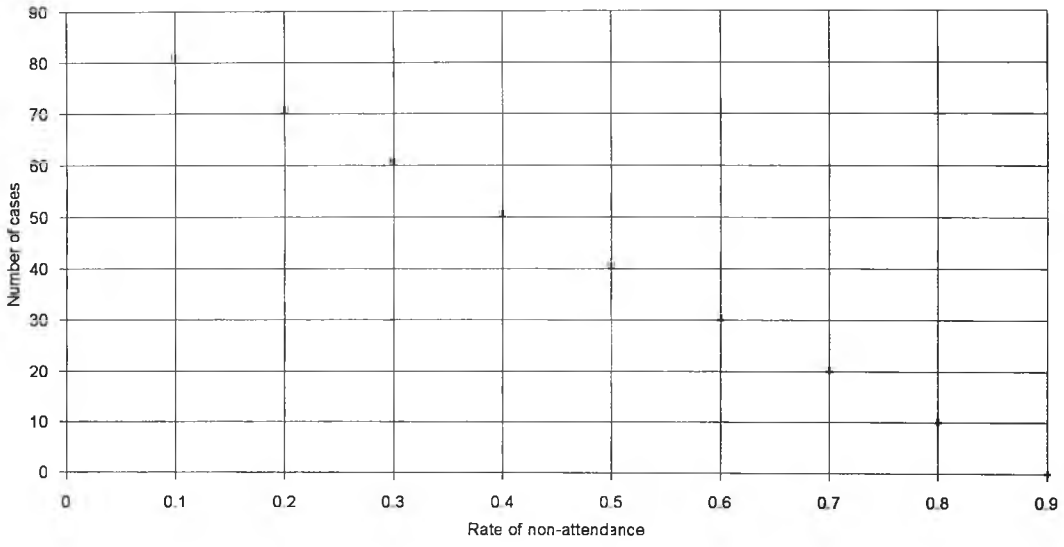


Figure 2.3.2.2.4

Effect of changing DNA rates on number ofcancers in the unscreened population

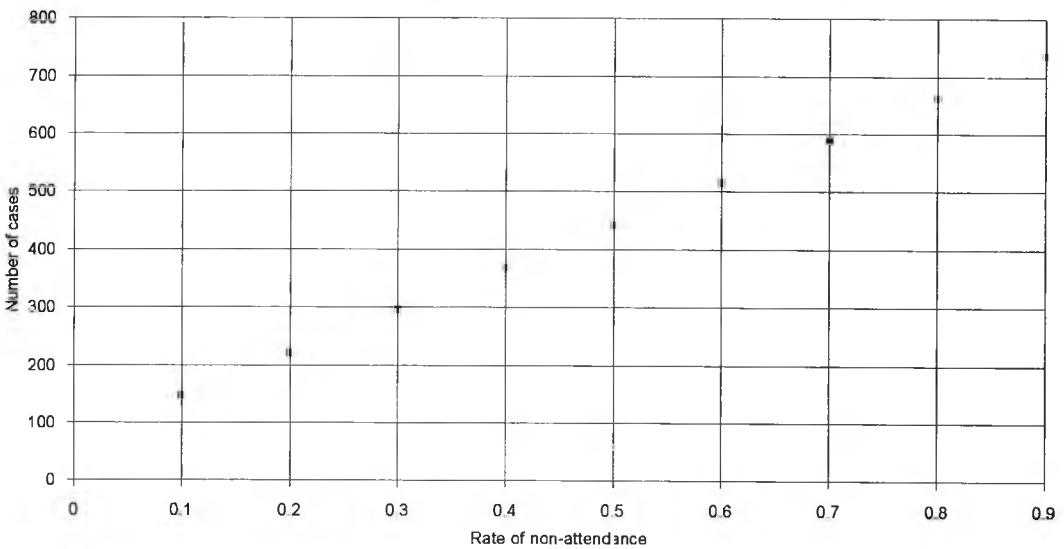


Figure 2.3.2.3.1

Effect of changing screening image quality on false positives at screening

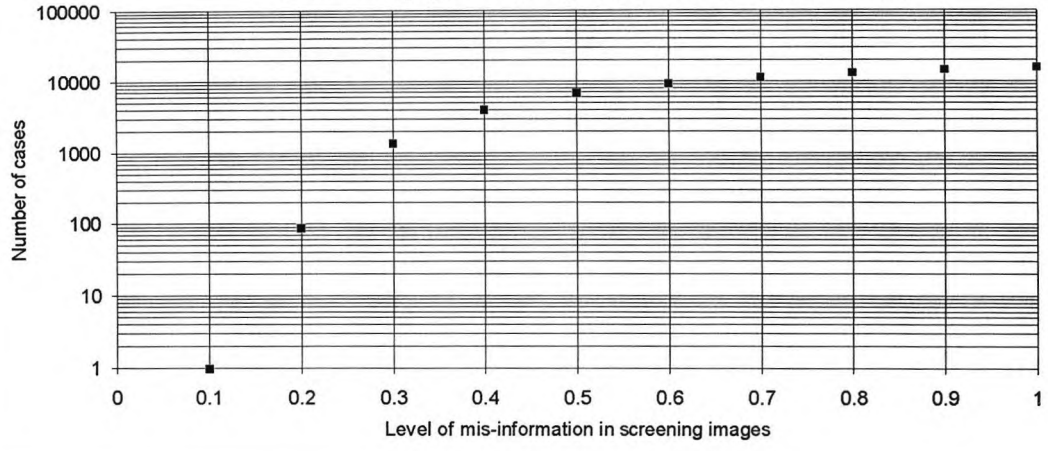


Figure 2.3.2.3.2

Effect of changing screening image quality on number of cancers detected

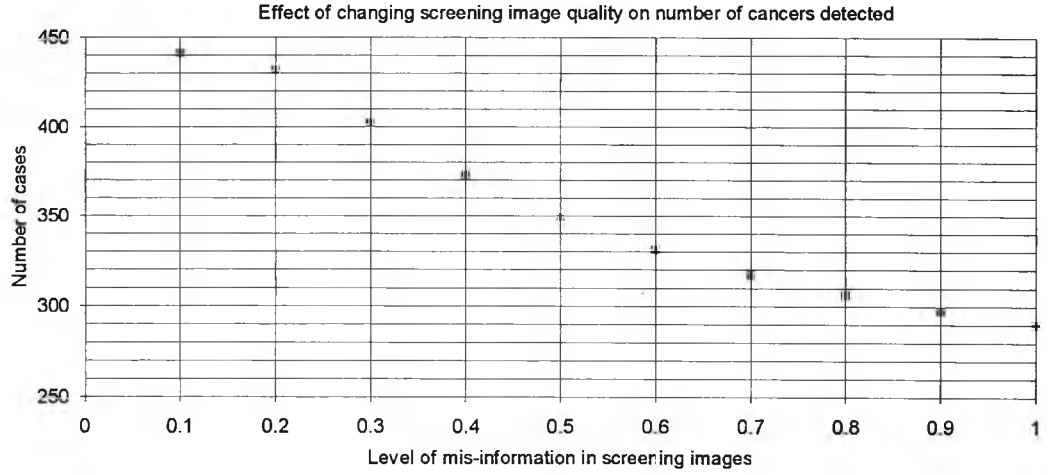


Figure 2.3.2.3.3

Effect of changing screening image quality on number of false negatives

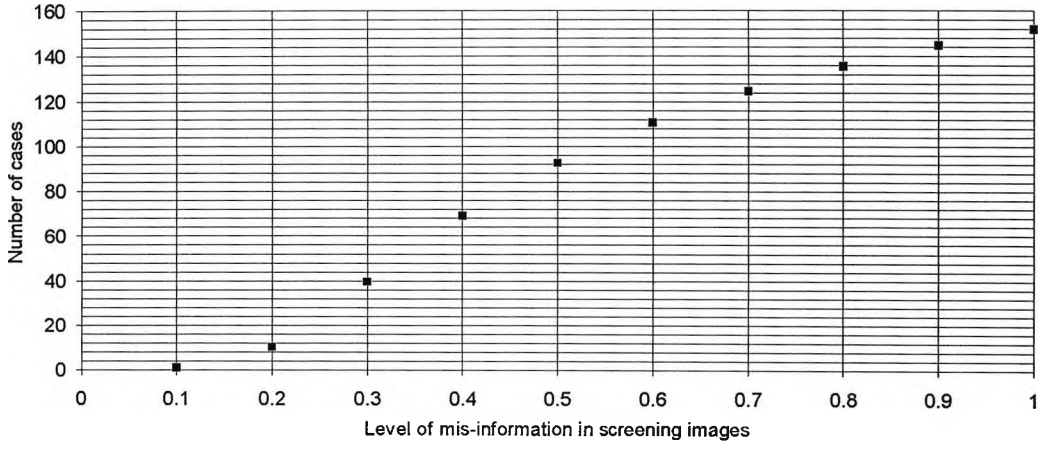


Figure 2.3.2.3.4

Effect of changing screening image quality on number of cancers in the unscreened population

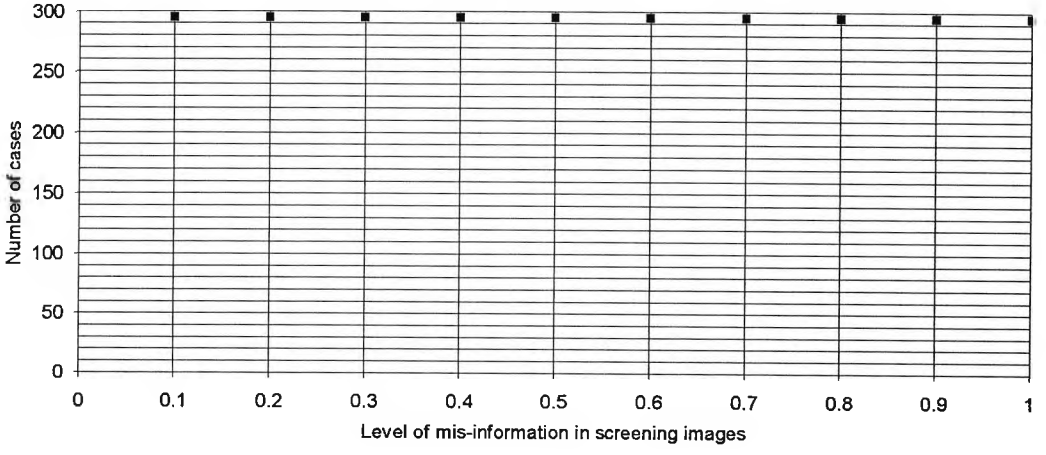


Figure 2.3.2.4.1

Effect of changing screening decision point on false positives at screening

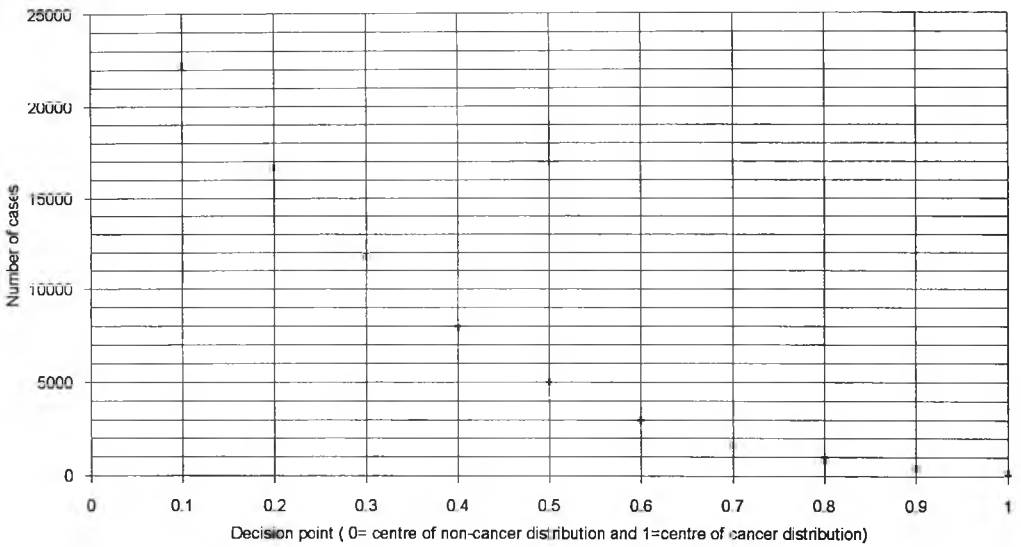


Figure 2.3.2.4.2

Effect of changing screening decision point on number of cancers detected

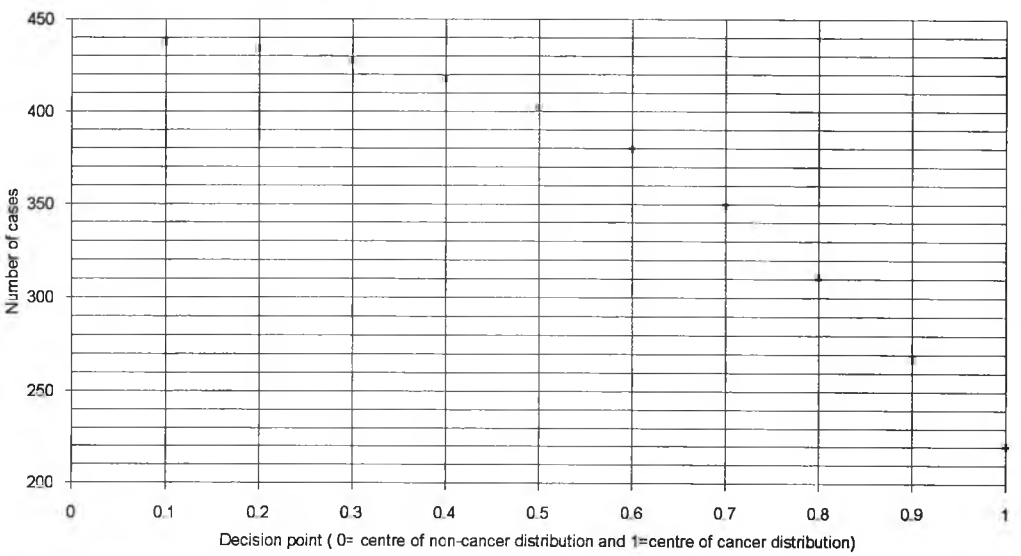


Figure 2.3.2.4.3

Effect of changing screening decision point on number of false negatives

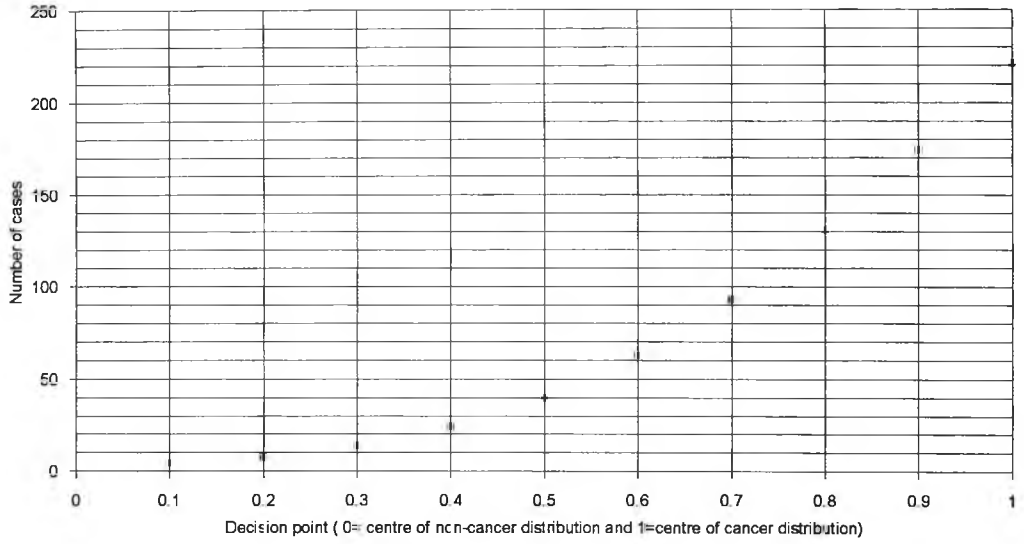


Figure 2.3.2.4.4

Effect of changing screening decision point on number of cancers in the unscreened population

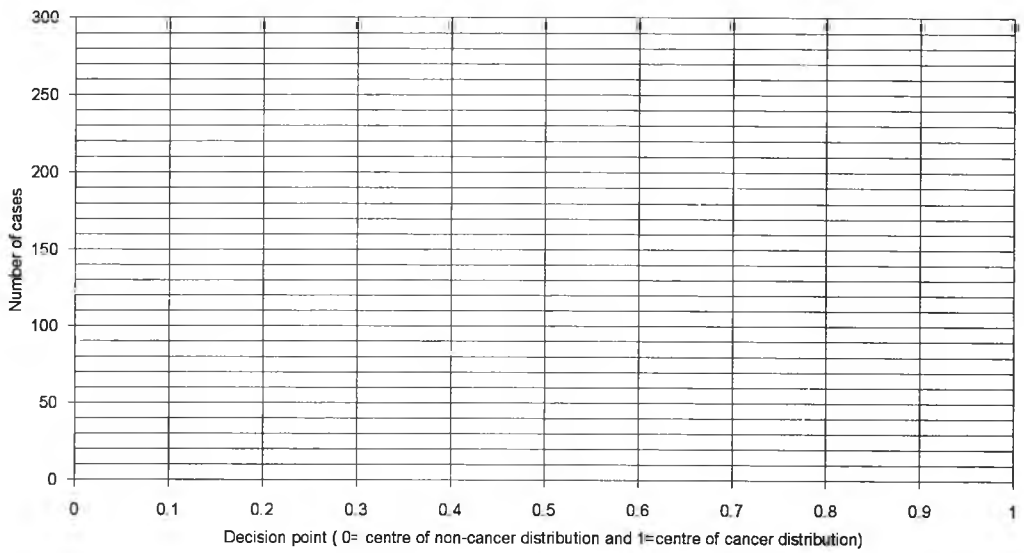


Figure 2.3.2.5.1

Effect of changing assessment image quality on false positives at screening

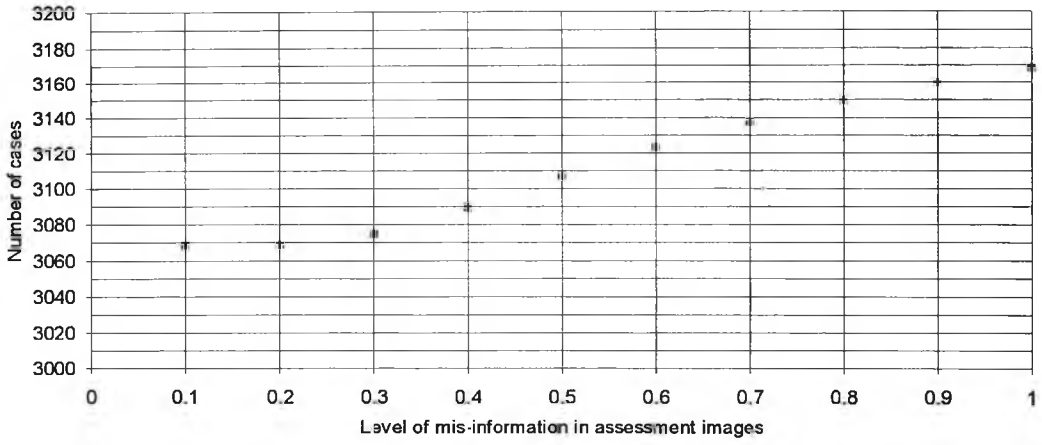


Figure 2.3.2.5.2

Effect of changing assessment image quality on number of cancers detected

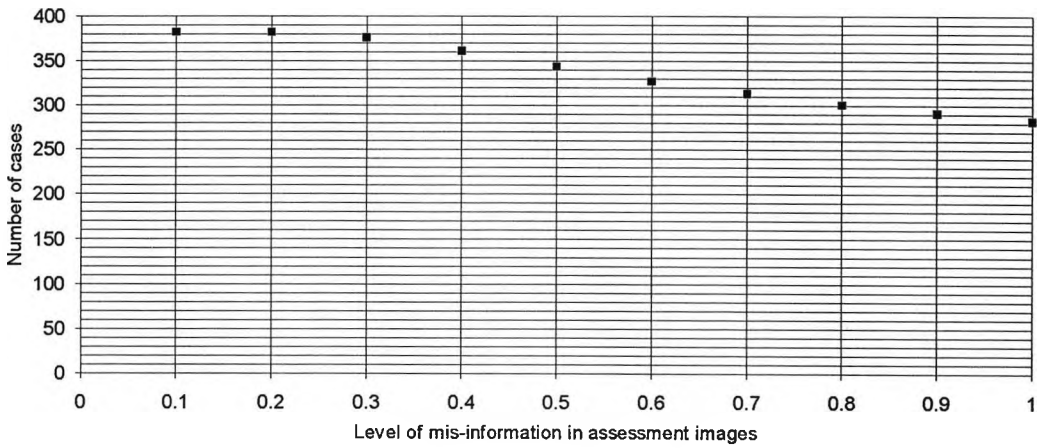


Figure 2.3.2.5.3
Effect of changing assessment image quality on number of false negatives

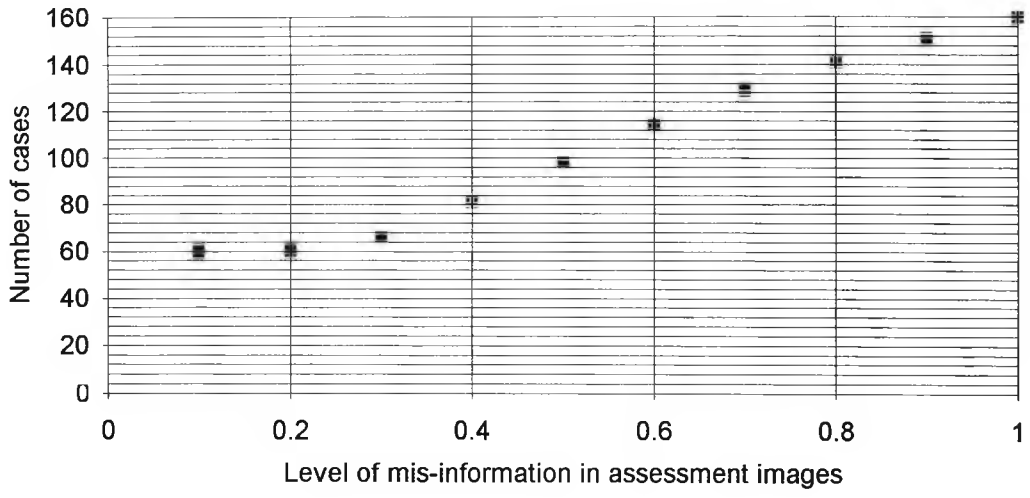


Figure 2.3.2.5.4
Effect of changing assessment image quality on number of cancers in the unscreened population

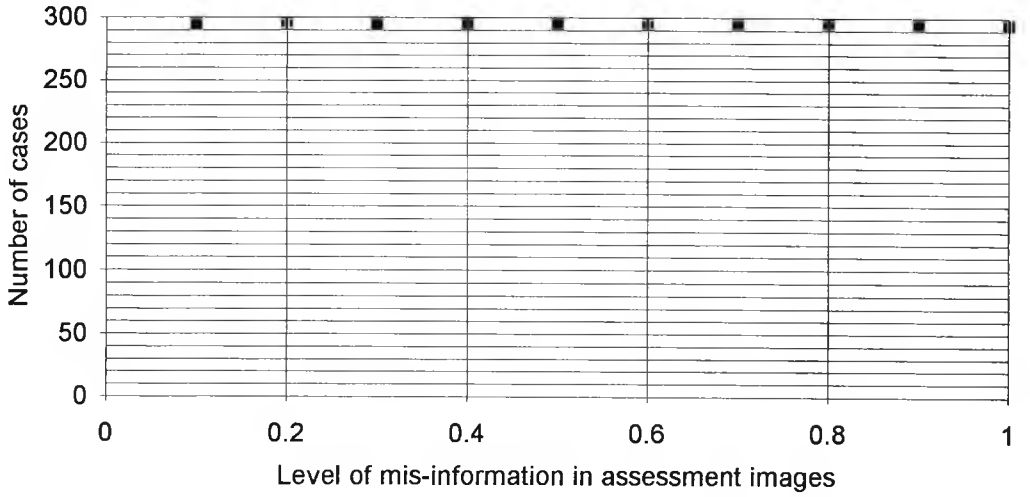


Figure 2.3.2.6.1

Effect of changing assessment decision point on false positives at screening

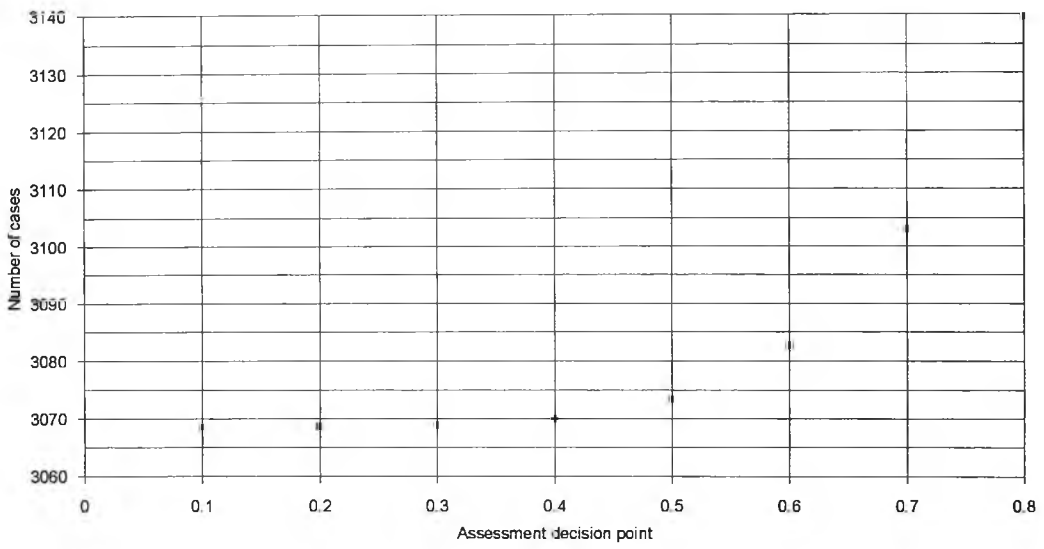


Figure 2.3.2.6.2

Effect of changing assessment decision point on number of cancers detected

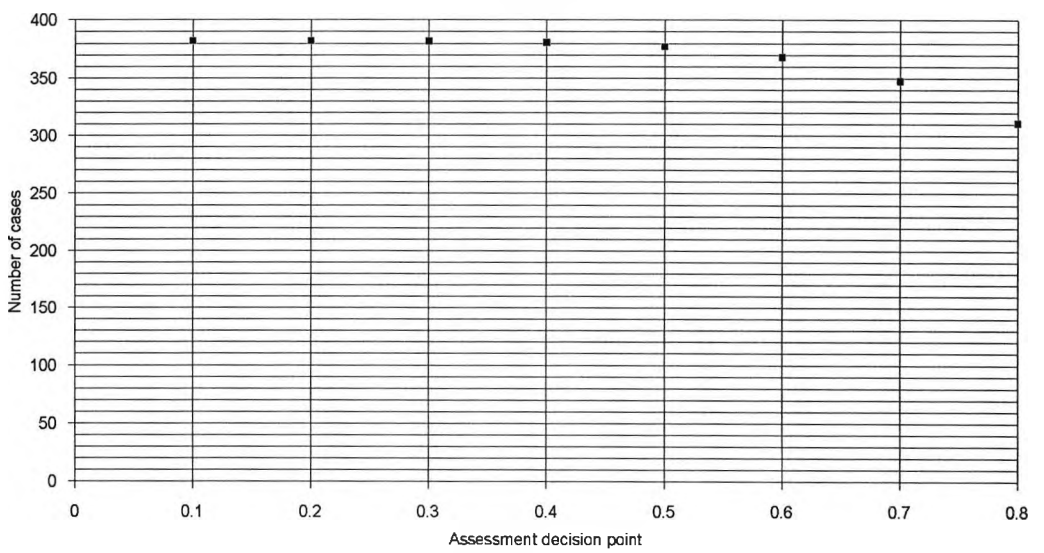


Figure 2.3.2.6.3

Effect of changing assessment decision point on number of false negatives

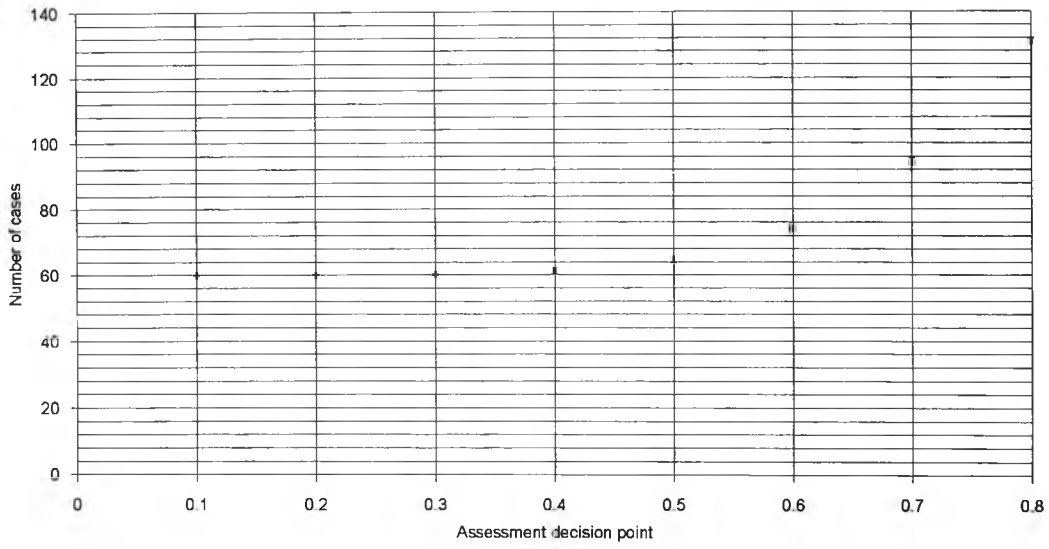
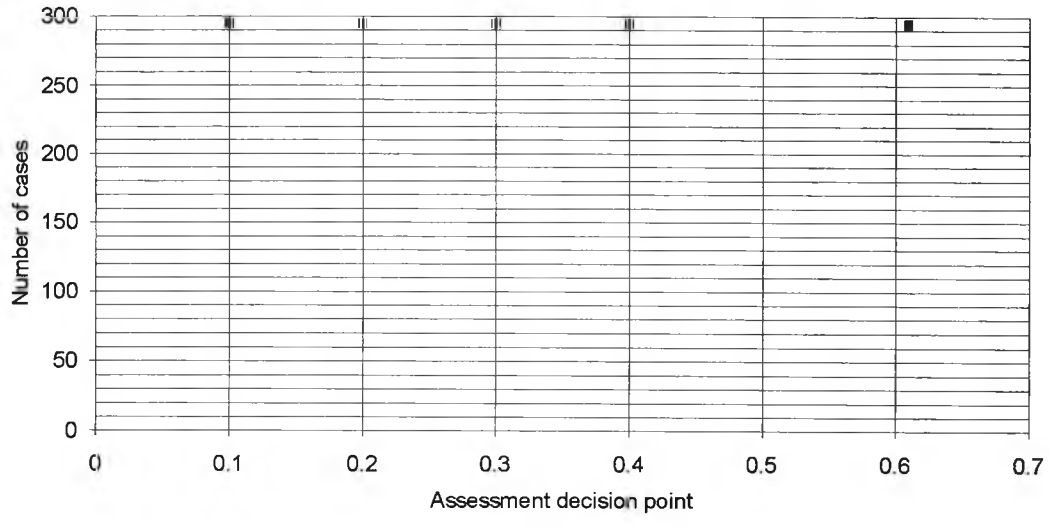


Figure 2.3.2.6.4

Effect of changing assessment decision point on number of cancers in the unscreened population



2.4 Desirable Technical Quality Requirements of Screening Equipment

It has been established, using the model of screening that the image quality has a substantial effect on the performance of the screening programme, therefore it is important to ensure that the image quality does not deteriorate over a period of time. The first stage is to identify exactly what is meant by a good piece of screening equipment. Having done that, it is necessary to ensure that all new equipment meets these standards and furthermore, that the equipment remains at that standard during its working life.

The two main requirements for radiographic breast screening are that the image produced should be of high enough quality to enable abnormalities to be detected, and that since the examination is being performed on a population of "healthy" women, the radiation dose should be minimised. These two requirements are closely related and often conflict. In addition, the equipment must be safe electrically, mechanically and in terms of radiation dose to any part of the body other than the breast.

There are two types of structure which are used to recognise a lesion, the presence of micro calcifications, usually calcium hydroxyappotite, which is highly attenuating to X-rays and shows up on the film a light speck and the solid mass itself which has a similar attenuation to the glandular structures in the breast and can therefore be distinguished from fat but not from glandular tissue, this is usually evident as a change in the breast architecture.

The imaging chain and factors which have to be taken into consideration at each stage is as follows:

1 X-ray production
- selection of spectrum,
- target material,
- filtration,
- tube voltage

2 Production of image by absorption in the breast
- compression,
- film to focus distance,
- focal spot size

3 Conversion of intrinsic X-ray image to latent image on film
- selection of film,
- selection of intensifying screen,
- automatic exposure devices

4 Processing of the film
- processing temperature,
- storage of film prior to processing,
- sensitometry,
- extended processing cycles

5 Reading of the film
- viewing conditions

2.4.1 X-ray production

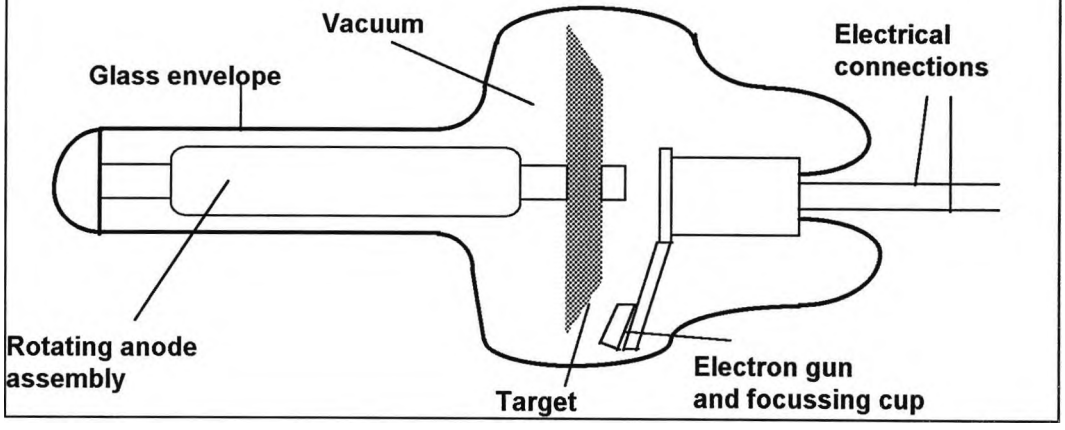
X-rays are produced when a high energy electron strikes a solid target [102], the target is usually made out of Tungsten in conventional radiography or Molybdenum in mammography. The electrons emerge from an electron gun which is a small hot filament at one end of an evacuated tube (figure 2.4.1.1) and are made to accelerate by applying a high voltage between the filament (the cathode) and the target (the anode). Because of the low voltages used across the X-ray tube in mammography, the filament is positioned closer to the anode than in normal X-ray tubes in order to reduce the influence of the space charge effect [103]. The intensity of the X-rays as a function of energy is called the spectrum and consists of two parts, the Bremsstrahlung which is a smooth function which cuts off at a maximum energy corresponding to the tube voltage, and the k lines which occur at distinct energies which are characteristic of the target material, a typical spectrum is shown in figure 2.4.1.2. The spectrum is further modified by any material which the beam passes through before impinging on the patient. Metallic filters are used in order to modify the spectrum in some desired way.

If a single photon energy were to be selected, a compromise must be made between the need to keep the radiation dose low which demands a high photon energy, and image quality, particularly image contrast, which demands a low photon energy. Beaman and Lillicrap [104,105] have produced a set of curves of the signal to noise ratio, which can be used as a measure of the image quality, as a function of photon energy for a range

of breast thicknesses when trying to detect 100 micron calcifications for a fixed dose. The energy at which the curve peaks increases with increasing breast thickness, the graphs taken from reference [106] are shown in figure 2.4.1.3. Bands of photon energies selected to give optimum signal to noise ratio by this method are;

2 cm breast 14-18 keV
4 cm breast 17-21 keV
6 cm breast 19-23 keV
8 cm breast 20.5- 23.5 keV

Figure 2.4.1.1 Rotating anode tube



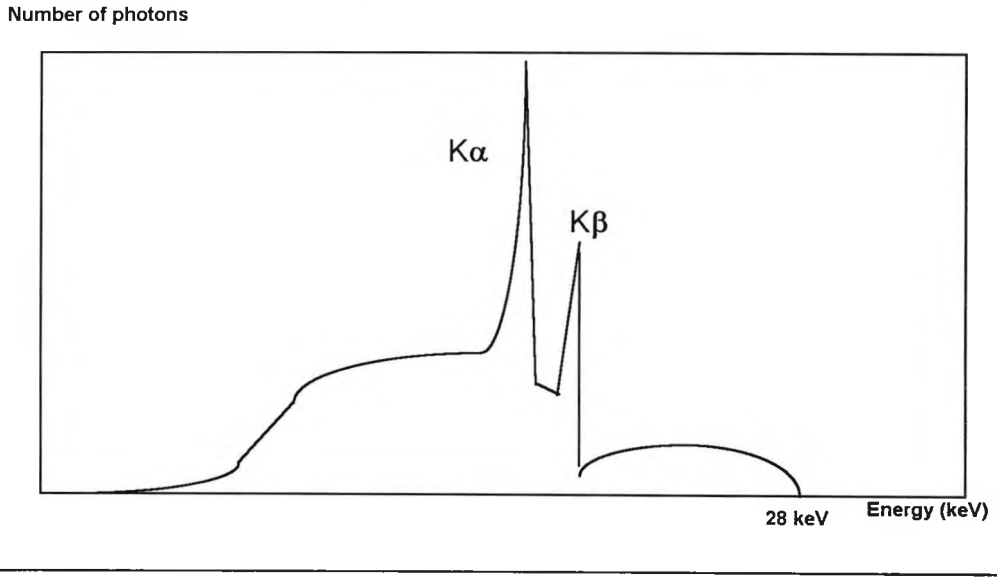
In order to change the spectrum from that originally emitted to one which predominantly contains photons in the desired energy band, the beam is made to pass through a metallic k-edge filter. When a k-edge filter is used, the X-ray spectrum above the k-edge energy is effectively chopped, giving an upper limit to the energies in the beam. The energy absorbed is re-emitted at the k-edge energy of the filter. Materials with a k-edge in the range of interest which could be used as filters are:

Molybdenum	k-edge 20.0 keV
Rhodium	k-edge 23.3 keV
Palladium	k-edge 24.3 keV

In addition to the Bremsstrahlung, the target material produces characteristic radiation peaks at its own k-edge energies, which in the case of Molybdenum is 18.5 (K_α) and 20.0 keV (K_β) producing strong peaks in the desired energy band.

The use of a molybdenum target with a 30μm molybdenum filter and a low tube voltage produces a spectrum which gives good or optimum signal to noise ratios for all but the largest breasts. The use of other filters is being explored by some manufacturers but currently a molybdenum anode with a molybdenum filter is the standard for mammography.

Figure 2.4.1.2
X-ray spectrum from a Mo anode
curve a) with no filtration
curve b) with a 30 micron Mo filter



2.4.2 Image production by absorption in the breast.

As in any X-ray image production, the image is produced by differential absorption of X-rays by different tissues. It is because the breast tissues are very similar in chemical composition that such low photon energies are required to distinguish between the tissues. At photon energies around 60 keV which might be used for a normal chest X-ray, calcification, muscle, glandular tissue and fatty tissue have such similar attenuation coefficients that little or no structure can be seen on an X-ray image. With a photon energy of 20 keV, there is a factor of 2 between the attenuation coefficient of fibrous tissue and carcinoma and the attenuation coefficient of fat [107]; this gives a contrast of around 0.01 (or 1%) for a 1 mm fibre in a 4.5 cm thick breast. Calcifications, being much denser and of different chemical composition at a size of 0.1 mm give a contrast of around 0.1 (or10%) at the same photon energy [107]. This data is however based on a very limited number of samples.

Compression is applied to the breast in mammography in order to reduce the dose, restrict movement and therefore increase image sharpness, and also to enable more uniform optical density to be achieved so that the whole of the breast is able to be visualised. An uncompressed breast would be underexposed in the thicker parts and overexposed in the thinner parts.

The size of the focal spot, typically a nominal value of 0.4mm for broad focus and 0.15mm for fine focus, is important in limiting the geometric unsharpness, the smaller the focal spot, the sharper the image. However, reducing the size of the focal spot limits the tube current and the heat able to be dissipated since it is all concentrated in a much smaller area; this in turn decreases the output and increases the exposure time which accentuates movement unsharpness.

In an ideal situation, the distance between the focus and the film would be as large as possible in order to reduce geometric distortion (this occurs because structures at the top of the breast are magnified more than those at the bottom). The penalty for doing this is that the radiation intensity falls off with distance squared and the exposure time required therefore increases and movement unsharpness becomes a problem. In practice, the distance between the focus and the film (FFD) is 60 cm and the radiologist has to make allowances for distortion when reading the films.

2.4.3 Production of the latent image on the film

Film emulsions are far more sensitive to light than to X-rays and as a means of dose reduction, it is routine to use phosphorescent screens which are placed in contact with the film, they absorb a large proportion of the radiation (around 90%) and convert it to light which then exposes the film producing the latent image. Gadolinium Oxysulphide is the most suitable phosphor for mammography because it is more sensitive to low energy X-rays than Calcium Tungstate which is used at higher energies [108,109]. The X-ray to light conversion efficiency is around 15% as compared to 3.5% for calcium tungstate. This leads to a substantial reduction in radiation dose to the patient [110].

Mammography film is unusual because it has emulsion on only one side and much smaller light sensitive crystals in the emulsion than is usual in modern medical X-ray film. This is to reduce image unsharpness, a single emulsion means that crossover is eliminated and small grains reduce the film unsharpness; it does however decrease the speed of the film and therefore increases the dose given.

The unsharpness of the film-screen combinations used is usually quantum limited rather than limited by grain size, the number of photons used to produce the image is so small that statistical fluctuations in the X-ray field produce more noise on the image than any other source. When dealing with counting statistics, it is usual to assume that the numbers involved form part of a Poisson distribution, because the processes involved are probabilistic. When this assumption is made, for a signal of n photons, the standard deviation is $\sqrt{n}$, [111]. In order for variations in the number of photons reaching the film to be statistically significant, due to a real variation in attenuation not just random fluctuations, the difference between the number of photons reaching one part of the film and its neighbouring part must be large enough to give a t-value of greater than 1.96 (p=0.05) The t-value can be calculated in the following manner

$$t = \frac{|n_1 - n_2|}{\sqrt{\sigma_1^{\,2} + \sigma_2^{\,2}}}$$ Equation 2.10

We know that

$\sigma_1 = \sqrt{n_1}$ and

$\sigma_2 = \sqrt{n_2}$

so we get $t = \frac{n_1 - n_2}{\sqrt{n_1 + n_2}}$ Equation 2.11

So if the value 1.96 is substituted for t, the value of n_1 required to produce a statistically significant result for a given intrinsic contrast level can be calculated. Details of the calculations appear in appendix D, a summary of these results are in table 2.4.3. As the dose goes down, the number of photons goes down in proportion, and, as can be seen from the table, the intrinsic contrast of the object needs to be higher in order to produce a significant change in the resulting optical density. Consequently, near quantum limits, the dose cannot be reduced without degrading the image quality.

Intrinsic contrast of object	No of photons required
1%	76,448
5%	2,996
10%	730
50%	23

Table 2.4.3 The relationship between the subject contrast and the number of photons required to produce statistically significant density changes (p=0.05)

The automatic exposure device is of vital importance in ensuring that the appropriate amount of X-rays reach the film. Because the intrinsic contrast in the X-ray image is so low in breast imaging, film screen combinations are selected which greatly amplify the contrast. A typical characteristic curve, which is a plot of optical density against log relative exposure is sigmoid in shape with a very steep gradient between the toe and shoulder regions. In order to produce images on the film which are neither over nor under-exposed, it is necessary to operate near the centre of this steep "straight-line" portion of the curve. There is a very narrow band of exposure values which satisfy this condition and accurate sensing of the radiation emerging from the back of the film cassette is necessary. In appendix C there is an example of a characteristic curve and an explanation of the relationship between its parts.

The automatic exposure controls in modern mammography units not only measure the integrated exposure and terminate the exposure when a pre-set amount of radiation has been detected but also measure the exposure

rate which gives an indication of the attenuation which the beam has undergone. A highly attenuated beam has a large proportion of high energy photons because the low energy photons are more strongly attenuated in their passage through the tissue (an effect known as beam-hardening) and the film-screen combination is less sensitive to higher energy radiation therefore slightly more radiation than the pre-set amount will be required. The X-ray units also monitor the tube voltage, this affects the X-ray spectrum and similar corrections are needed to compensate for the changes in photon energy.

2.4.4 Processing the film

In order to minimise the loss of radiographic information, the film should be processed immediately after it has been exposed. On mobile units far away from base, this is not possible and films are processed several hours, or sometimes longer, after they have been exposed. It is inevitable then that there will be some fading of the latent image [64], this is kept to a minimum by storing the exposed but unprocessed films in cool, dry conditions. In order to produce greatest contrast amplification, mammography films are processed at higher temperatures than normal, 35°C is typical, and are processed for a longer time than usual, 3 minutes in the processor rather than the usual 90 seconds. Murray et al [64] recommend 3 minutes transit time in the processor and a developer temperature of 36°C. In effect, they are deliberately overdeveloped or "push-processed" and because other aspects of the system such as the AEC can be so finely controlled, the films still span the correct optical density range. Again, because every link in the imaging chain is so delicately balanced, the processor has to be dedicated to mammography films and rigorous daily checks kept on the processing. Thus at the first sign of change in the processor, corrective action can be taken.

2.4.5 Reading of the film

The reading room should have a low ambient light level and the viewing box should be brighter than usual because the darker areas of the film will be darker than usual due to the push-processing. This can cause problems with glare from areas which are not covered by film. Arrangements need to be made to ensure that these areas can be masked off or the light turned off. In screening, roller-viewers are often used, a whole day's worth of films can be loaded side by side, thus

reducing glare. These viewers have the facility to control brightness and to select which areas of the screen are turned on. This helps to reduce fatigue in the observer both by providing favourable viewing conditions and by removing the need to take several hundred films out of their packets.

2.4.6 Assessment Equipment

At the assessment stage, the women being examined have already undergone routine screening and have been found either to have some kind of abnormality which needs to be definitively diagnosed, or to have suspicious areas on the film for which further views or special techniques are required in order to produce a firm classification of the films as normal or in need of biopsy. At this stage, image quality is of primary importance and the radiation dose may be allowed to increase beyond levels which would be acceptable in screening in order to produce accurate diagnoses. In addition to radiographic techniques, the complementary techniques of clinical examination, and ultrasound examination are currently used. While they are unsuitable for initial screening due to low specificity, once an abnormality has been identified they are useful to confirm or refute the initial diagnosis. It is possible that CT, MRI or Doppler ultrasound could also fulfil this role in the future.

In addition to the requirements for screening equipment, which have already been discussed, the option of using a smaller focal spot is desirable, as is the facility to produce magnified views of the breast radiographically. It is helpful to have some form of localisation facility either a stereotactic device or a perforated compression plate in order to take aspirate samples for cytology from lesions although skilled operators are able to do this under ultrasound guidance [112].

2.5 Details of Quality Assessment

In this section, the tests which were performed on the equipment in the breast screening programme will be described, these are the tests recommended by the Institute of Physical Sciences in Medicine [4] and adapted to suit the working practices in North East Thames [113]. The limitations on the accuracy of the tests will be discussed in detail in chapter 3.

2.5.1 X-ray equipment

Tube voltage

The tube voltage should be calibrated to an accuracy of ± 1kV. This can be checked using an electronic meter which is designed specifically for mammographic purposes. The meter is placed so that its sensitive area is within the beam and an exposure is made. The meter used is insensitive to positioning so long as the marked sensitive area is within the defined beam area. There is some exposure dependence, the meter requires a reasonably large exposure, 20 mAs or more, in order to function correctly.

Radiation Output

Consistency, variation with kV and variation with mAs are measured using an ion chamber with a good low energy response, which must be placed so that the whole of the ion chamber lies within the beam. The consistency should be better than 10% of the mean value, the output should be approximately proportional to kV^2 and the output per mAs should be independent of mAs (therefore this should also be within 10% of the mean value to maintain consistency). Positioning of the ionisation chamber is critical to absolute values but not to relationships (e.g. the output is theoretically proportional to kV^2) however, IPSM 59 [4] suggests that a distance of 40 to 50 cm from the tube focus to the ion chamber is suitable and that the ion chamber should be at least 10 cm above the cassette table in order to minimise backscatter.

Exposure Time

This is measured simultaneously with output. If the tube current is constant throughout the exposure which is the case with a correctly functioning medium frequency generator, its value can be inferred from the ratio of mAs to measured exposure time. For short exposure time (< 200ms) the consistency should be within 15% and for long exposure times, a figure of 10% has been suggested.

Half Value Layer

The half value layer (HVL) is a measurement of the amount of filtration which is needed to attenuate radiation from the primary beam by a factor of two. A range of thicknesses of aluminium are placed in the beam and the transmitted output is measured at each thickness, graphical interpolation is used to estimate the thickness which reduces the radiation intensity to half of its original value. The HVL is an indicator of beam quality and therefore varies both with tube voltage and with filtration, it doubles as a safety check to determine if sufficient filtration is present and is also required for dose calculations. Narrow beam geometry should be used to minimise the influence of scattered radiation on the results. The HVL should be greater than 0.3mm Al to ensure that there is adequate filtration in place, there is no upper limit on HVL, although if the filtration is too great, the output is reduced so much that movement unsharpness due to long exposure times may become a problem.

Automatic Exposure Control

The termination of an exposure should produce the same optical density on the film regardless of the kV selected or the size of the attenuator. Perspex (poly methyl methacrylate) is used as the attenuator for AEC testing. The consistency of mAs and optical density for a fixed kV and density setting with the AEC on zero, which constitutes the default setting, and the consistency of optical density with varying kV and with varying attenuation is tested and should be within 10% of the default mean in all cases.

Compression

This test is primarily a safety test and checks that an upper limit of 200 N is not exceeded [1]. It is also important that the force produced is maintained and that there is no backlash of the system. The measurement is carried out using a force balance which is able to adjust the angle of the sensitive component to align with the compression plate, the balance scale has a separate pointer which records the upper force registered on the display.

Beam Alignment

This is a safety check to ensure that radiation is emerging only where it is expected to, and is aligned not only with the light beam but also with the film. In order to ensure that as much of the breast is imaged as possible, the requirements for the chest wall (0 to 3mm) are much more stringent

than for the other three borders (±5mm). The test is performed by placing a loaded cassette in the bucky and then placing a larger cassette (or two small cassettes) on top of the breast table so that there is an overhang of a couple of centimetres. Radio-opaque markers are placed on top of this to mark the edges of the light beam and an exposure is made. By comparing the films and measuring the positions of the markers on them, the relative position of the light beam, X-ray beam and film can be calculated.

Focal Spot Size

This cannot be adjusted since it depends on the physical arrangement of components within the X-ray tube [114]. It is claimed that the focal spot size should not change. On one unit in Scotland this has not been the case [115], however the significance of this observation is uncertain. If material is being evaporated from the anode, the shape of the focal spot might be expected to change, it is also possible that the focal spot will change when the bias voltage on the focus cup changes, however, not all tubes have such a control. The resolution of the system depends very strongly on the focal spot size [116]. In order to make this measurement, a pinhole or slit may be used to act as a camera producing a magnified image of the focal spot on a film. The alternative method is to calculate the focal spot size from the resolution limit measured using a star pattern. It is necessary to ensure that the axis along which the measurements are made is the same as the beam axis specified by the tube manufacturer [117] otherwise the results will be incorrect due to the line focus effect [118]. A special jig is required in order to achieve this in mammography because, unlike conventional diagnostic radiography, the central axis is not usually vertical [117].

Mean Glandular Dose (derived)

The calculation of mean glandular dose is based on the tube output, the mAs recorded on automatic exposure control and the beam quality (as indicated by half value layer measurements). Backscatter factors and air kerma to dose conversion factors are based on Monte Carlo Calculations [119,120] and are related to the measured half value layer. The calculated mean glandular dose is that for a standard breast, and should represent the median dose for a population. There is no lower limit on mean glandular dose, IPSM 59 [4] gives an upper limit of 3 mGy per exposure and Pritchard [3] gives a limit of 5mGy per view, this was subsequently

revised downwards by the national breast screening programme to 2mGy per film for a standard breast [121].

Image Quality film

The test object gives a record of how the whole system performed on a certain day. There are some subjective features, such as blocks of simulated micro calcifications and some semi-quantitative features such as large low contrast discs, smaller higher contrast discs and limiting resolution test patterns. There is also a stepwedge which covers the low optical density end of the characteristic curve of the film and screen in response to X-rays. Pritchard [3] set lower limits on the number of each type of object which should be visible if the system is producing acceptable image quality. Should the image quality film be sub-standard, further investigations then need to be undertaken to ascertain the cause.

At acceptance, the test object is very useful in setting up the AEC to give an optimum density level. It should be remembered that the test object does not exactly reflect clinical performance, and that the visual acuity of the physicist or engineer involved in the setting-up process may be very different to the radiologist who will use it routinely, and further fine adjustments may need to be made once the system has been used clinically for some time.

X-ray Field Non-uniformity

Perpendicular to the anode-cathode axis the radiation field should be uniform to within 10% so that structures at the edges of the film are not under exposed. In the anode cathode direction, the anode heel effect is deliberately used to produce increased intensity at the chest wall edge in order to penetrate the denser structures which are normally found there. It is measured by exposing a film and measuring the optical density at various points on the film. Sensitometry can be used to make the conversion between optical density and exposure.

Magnification

This depends solely on geometric factors and therefore only has to be measured at acceptance. The check is intended to ensure that the quoted magnification is achieved in practice. A test tool with radio-opaque markers at known positions is exposed and the size of the image on the film is measured and compared to the known dimensions.

Secondary Radiation Grid

The grid is used to reduce the amount of scattered radiation reaching the film, it also increases the amount of incident radiation required in order to produce the chosen optical density on the film. This increase in dose should be less than a factor of three to ensure that a reasonably low dose to the patient is achieved, and is measured by producing sensitometric curves both with and without the grid in position and comparing the exposure required to produce a density of 1 + fog in both cases.

Radiation Leakage

This is purely a safety consideration, the radiation should emerge only from the collimator so as to avoid irradiating either staff or parts of the patient which are not in the imaging field. One would not expect to find any measurable level of leakage, there are statutory limits that the leakage radiation should be no more than 1mGy in 1 hour at a distance of 1m from the focus averaged over an area of 100 cm^2. [122]

Table Transmission

A safety consideration, the table should be constructed of sufficiently dense material that none of the incident radiation beam emerges from the other side of the table. There is currently no legislation on permissible limits of radiation emerging from the far side of the table, but because the aim is to prevent parts of the patient which are not within the primary beam from being irradiated, the limit for radiation leakage from the tube housing would appear to be a sensible one to adopt.

Earth Bonding Test

This is a safety test to ensure that all of the components of the equipment which can conduct electricity are earthed and not liable to produce an electric shock in the event of a breakdown of the high tension insulation. All metal components must have an impedance of less than 0.1Ω with respect to ground. This is normally only tested at acceptance [123].

2.5.2 Processing [113]

Sensitometry

This is done daily and the results are plotted on a control chart, action limits are set and corrective action is taken when these limits are exceeded. Full details are in appendix C. It monitors the way in which the

processor is varying and any changes which may occur in the film, for example, when a new batch is delivered or if the film is being stored under conditions which are less than ideal.

Cassette Sensitivity

The thickness of the phosphor layer is not always identical from one cassette to another, this leads to different relative speeds and unwanted changes in optical density. All new cassettes are checked and should be rechecked at six monthly intervals to ensure that they have not deteriorated.

Film-Screen Contact

Poor contact between the film and screen causes unsharpness of the image. This must be periodically tested but would not be expected to change unless mechanical damage (such as scratches on the phosphor or bending) had occurred. The test is performed by placing a test tool, which contains small regularly spaced holes, on the table and making an exposure. In areas where the film screen contact is poor, the pattern becomes unsharp and dark areas are observed.

Safelight Handling Time

Poor darkroom conditions cause fogging yet are imperceptible to the naked eye. This must be tested periodically to ensure that the filters used on the lights have not degenerated. The simplest method to test this is to print a sensitometry stepwedge on several films in complete darkness. The first film is processed immediately, subsequent films are exposed for progressively longer intervals of time to the safelight. The optical density at the mid grey level, which is the most sensitive to fog, is then compared to see if the safelight has caused any fogging.

Retake Analysis

This procedure is performed routinely on all films which have to be retaken. Its function is to identify areas which need attention and action. Usually, the reject films are stored in a designated area and they are periodically categorised.

Image Quality Film

As discussed in section 2.5.1, the image quality film is used to monitor the performance of the imaging chain. If the image quality has been degraded by the performance of the processor, it will also be observed in

sensitometric testing as well as with the test film. The image quality film is also able to demonstrate mechanical processor faults such as scratches, pick-off or roller marks which may not be obvious in sensitometry

2.5.3 Viewers

Image Quality Film

Evaluating the image quality film is subjective and is affected by a range of factors including the viewer. If the image quality film appears to be poor but gives acceptable results on other viewing boxes then a problem with the first viewing box is indicated.

Light Levels

Ambient light levels within the room need to be low, 86lux or less and the viewer needs to be bright, 5500 lux or more, to cope with some of the high optical densities encountered [4]. A light meter is used to make these measurements.

Uniformity Across Viewer

It is also important that the whole of the image is uniformly illuminated. This can be tested with a light meter or with a hand held densitometer placed on the viewer at various positions. There should be no more than 10% variation away from the mean across the viewer.

2.6 Definitions of Required Quality - Literature review

There is only a small amount of literature which is directly relevant to quality assurance in a medical setting. The statistical QA techniques in industry are well established and can be found in a range of statistical quality control textbooks [124]. The industrial methods are complicated by the fact that in radiography, safety issues are also pertinent; although achieving consistency is important, it is not the sole consideration. In terms of dose limitation, there are four sets of relevant regulations [125,126,127,128] plus a set of recommendations to the Royal College of Radiologists [129] the parts of which are relevant to mammography can be summarised as follows;

1 Suitable Precautions must be made to ensure that the region of interest of the person being screened is the only thing subjected to radiation (beam limiting, protective screen for the radiographer).

2 Records must be kept which allow a reasonable estimate of dose to be made.

3 Special filtration is required for mammography in order to keep the dose low, never less than 0.5 mm of Aluminium or 0.03 mm of Molybdenum.

4 Mammography should not be practised on women under 50 years of age. This is based on the recommendations in the Forrest Report [2] and should therefore not be taken as an absolute recommendation, the value of mammography for screening in the 40 to 49 age group is of uncertain value and is the subject of a number of research trials.

5 Previous films should be made available in order to avoid unnecessary examinations.

The UK Breast screening programme has developed its own QA guidelines [3] which address a range of issues. As far as physical measurements are concerned, it addresses what should be done but gives little indication of how or to what specification. It refers to IPSM 59 [4] as the appropriate source of such information, it does however specify in its Table 2 a maximum dose of 5mGy per view and some image quality targets based on the use of the Leeds TOR(MAX) phantom.

In accordance with the underlying philosophy of radiation protection, every precaution must be taken to ensure that the dose received is as low as possible consistent with the desired clinical result, a diagnostic quality radiograph, [130] and that the benefit of the examination must outweigh the risk.

The need for uniform high image quality, although desirable, is not of paramount importance in diagnostic radiology as it is in mammography. The quality of an image, because a human eye is the final receptor in the imaging chain, is highly subjective, but also very adaptable. One method which is commonly used to attempt to quantify this is that of using a test object containing series of details, and a systematic way of assigning a score to each image according to the number of objects which can be seen.

The European Quality Criteria for Diagnostic Radiographic Images [130] sets out the following criteria for good imaging performance;

The unit should be able to image round details down to 3 mm in diameter and micro calcifications down to 0.2mm, however, no contrast level is defined.

The dose quoted for the entrance surface dose for a 4.5 cm compressed breast using an anti-scatter grid is 7 mGy.

There are also a number of items listed under the heading "EXAMPLE OF GOOD RADIOGRAPHIC TECHNIQUE " which are more relevant to the purchasing of equipment.

1. Radiographic Device: specially dedicated equipment with Molybdenum anode
2. Focal Spot Size: less than or equal to 0.6 mm
3. Total Filtration: 0.03 mm Molybdenum or 0.5mm Al equivalent
4. Anti-scatter Grid: specially designed moving grid might be necessary
5. Film-Screen combination: Dedicated high-resolution with dedicated processing
6. Focus to Film Distance: greater than or equal to 60 cm
7. Radiographic Voltage: 25 - 35 kV
8. AEC: chamber selected- specially positioned
9. Exposure time: less than 2 s
10. Breast Compression: Should be applied to a level which the patient can tolerate.

These standards, although useful, do not specify the way in which some of the quantities are defined, for example, the focal spot size could be the size quoted by the equipment manufacturer or the size measured by one of a variety of techniques which may vary by as much as 50%.

Murray et al [64] recommend 3 minute processing time (46 Seconds in the developer) and a developer temperature of 36°C. IPSM 59 [4] gives no specifications, it simply refers back to those set by the manufacturer of the processor.

The world health organisation [38] specifies that the focus shall be 0.6mm or less, ideally between 0.3 and 0.5 mm. The X-ray unit should be one specifically designed for mammography with a focus to film distance in the range 55cm to 75cm and automatic exposure control should be available. The remainder of the document refers to the need for specially trained staff and controlled procedures.

Hessler et al [131,132] set out a method for quantitatively assessing mammography systems, they defined an image quality index which is a function of contrast, spatial resolution and the overall noise which comes from three sources, film granularity, quantum mottle and screen structure mottle. They also measure the surface entrance dose to the breast using TLD's so that the whole assessment can be done remotely by mailing the test object and TLD's to the unit who make the exposure and send the test film and TLD's to a fully equipped laboratory where the assessment will be performed. They found that the image quality does not correlate with the optical density of the film but that the contrast and noise do. They suggest no quantitative limits, instead, they take the statistical approach, looking for units which are significantly different to the main group and attempting to identify the reasons and make rectification.

There are also a set of recommendations from the ACPSEM [133] which apply in Australia and are firmly based upon IPSM 59 [4]

In a report on five years of equipment quality assurance, Karila [134] reports on measurements of contrast, film optical density, resolution, focal spot size, tube voltage, exposure time, HVL, linearity and reproducibility of exposure, surface and exit dose and average whole breast dose.

Chapter 3

X-ray Units

3.0 Introduction - Why the X-ray equipment is tested

The rationale for testing components of the imaging system is twofold, firstly, to guarantee the safety and fitness for use of the equipment and secondly, to ensure that standards of image production over a period of time are maintained and long-term drift is detected and suitable correction is made. A secondary benefit is to monitor ageing processes and if economically viable, to have planned component replacement rather than disruptive repairs when components finally do fail.

The safety testing needs, theoretically, to be done before each patient is examined, although in practice, if the equipment has been established as working correctly before the first patient, the correct performance during the examination of any one patient may be used to indicate that the equipment is fit for use with the next patient. This will not prevent catastrophic failures in routine use as these, by definition, are sudden and provide no prior warning. In practical terms, this means that the equipment should be safety tested when it is first turned on, either once a day or twice a day depending on whether the equipment has been turned off over the mid-day period.

Long term constancy testing, on the other hand, aims to detect trends which are so subtle that on a day-to-day basis any changes are hidden by the random noise of the measurements. Measurements done on a day to day basis are very useful in determining drift of the system as a whole, but supplementary measurements need to be done periodically in order to determine exactly which part of the system is responsible for the drift. The frequency of these measurements needs to be sufficient that changes do not go undetected but economically and logistically it is preferable that the measurements should be as far apart as possible. The frequency must be determined by the ability to detect a signal above the noise arising from the measurement system.

3.1 Determining the optimum frequency for testing system components.

The method used to identify the optimum frequency for testing of components is straightforward, taking the 16 X-ray units in the North East Thames region as sources of data, the changes in various parts of the system have been

monitored at four monthly intervals, or more frequently for certain parameters. A rate of drift in each parameter has been determined for each X-ray unit, by combining these together, it is possible to establish a range and standard deviation of drift with time. Derived values such as the output coefficient, which has been defined as the output per mAs per kV^2, have also been evaluated in the same manner. It may be possible that one or two of the measured quantities may be of particular value in detecting the deterioration of the system and some may be of no value whatsoever, it is desirable to know the predictive value of any particular measurement in order to establish a quality assurance programme which is as cost-effective as possible.

As was stated in the introduction, it is necessary to distinguish the signal, the drift in parameter values, from the noise, the random fluctuations which occur either in the X-ray unit or in the measuring technique.

3.2 Sources of noise and methods of determining their magnitude

For each parameter there are three sources of noise;

a Random fluctuations of the X-ray unit
b Random variations in the measuring equipment
c Variation due to the measurement method.

In order to determine the size of each of these contribution to the noise in the measurements, the following techniques have been used.

a Random fluctuations of the X-ray unit

The magnitude of these variations can be determined by long term observations of a system which shows no drift, statistical limits can then be set to estimate the noise due to random fluctuations.

b Equipment reproducibility

Long term, this should be routinely monitored by calibration against a standard, short term reproducibility can be determined by repeated measurements on a system which shows no drift throughout a day or repeated measurements over a period of time which is sufficiently small for negligible drift to have occurred.

c Set-up reproducibility

This can be determined by controlled variation of factors in the set-up and recording their effect on the measured quantity or where possible, using a measurement technique which ensures that certain factors are held constant (e.g. a single test cassette in AEC measurements). The likely size of such variations can be estimated and the result of the variations on the measured quantity calculated. If a relationship between the various quantities is known or can be found, the error estimate is easier.

Each of the measurements will be considered in turn and the sources and magnitudes of variations in measurements will be calculated.

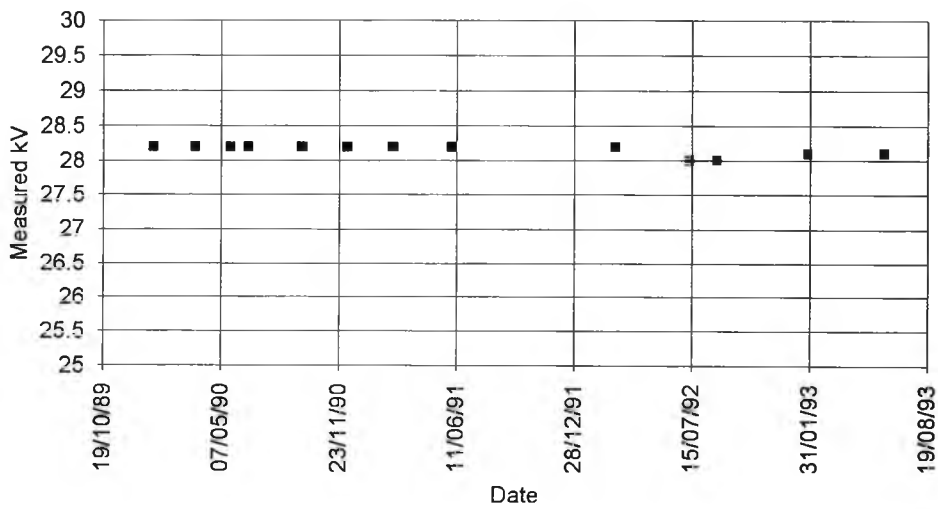
3.2.1 X-ray tube voltage

3.2.1.1 Variation due to normal statistical fluctuations

There may be some variations in the kV measured which are due to fluctuations of the X-ray unit and may be caused either by mains fluctuations or the fluctuations in the performance of the X-ray unit itself. The time series of tube voltage against time was plotted for several units and best fit straight lines calculated for each one. A unit which showed no long-term drift was chosen for the calculation (figure 3.2.1.1), this shows a standard deviation of 0.08 kV with a mean value of 28.15 kV giving a coefficient of variation of 0.3%.

Figure 3.2.1.1

Variation of kV in a system with no drift - Central and East London Mobile 1

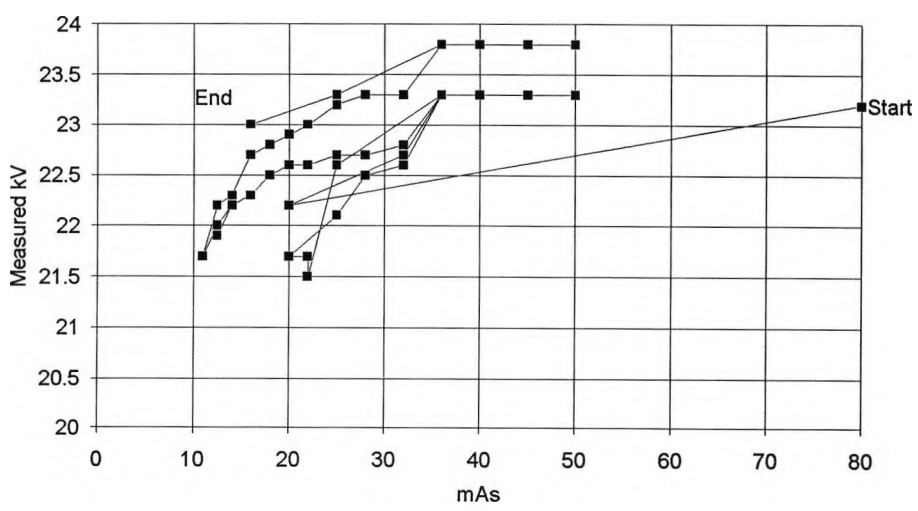


3.2.1.2 Variations due to the performance of the meter

It has been observed that under certain circumstances, the kV measurements are dependant on the functioning of the instrument. A series of kV measurements were made using the same tube voltage, 25 kV, the lowest which the X-ray unit was able to use. The mAs was varied between 11 mAs (the minimum which was able to register an exposure) and 50 mAs (by which point the measured kV had reached a plateau) and then back down again, allowing approximately thirty seconds for the X-ray tube to cool between successive exposures. Figure 3.2.1.2 shows not only the variation of measured kV with changing mAs but also the warming up of the meter over a period of time as the mAs goes from its lowest set value to its highest set value and back again several times. It can be seen that as the meter warms up, not only does the plateau value of kV increase, but also the meter becomes able to register an exposure at progressively lower mAs values. Ratings charts were not provided with the unit because the maximum mAs which could be set at any given kV was controlled by the system microprocessor, problems at low mAs had not been anticipated by the manufacturers.

Figure 3.2.1.2

Warm-up of kV meter



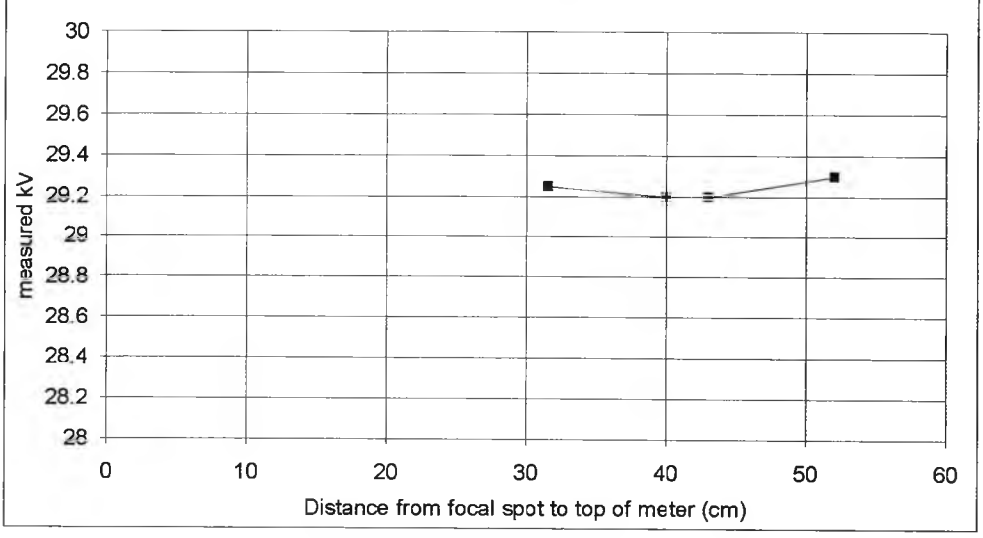
This effect was not reproduced at higher kV settings; for example, at 28 kV there was some reduction in the measured kV when the mAs was less than 5 mAs but no change of value at the plateau measured over time. The effect could have been due to focal spot bloom but is believed

to be due to inaccuracies of the kV meter at very low exposure rates; evidence in support of this hypothesis is that a failure to measure kV correctly had also been observed when measurements were made on the Philips mammography unit. This unit had a much lowere output per mAs than the Siemens units and the meter needed to be raised well above the breast table, taking advantage of the inverse square law to increase intensity, in order to produce an accurate reading. During QA measurements, 40 mAs was used to ensure that the meter was always functioning in the saturation, or plateau, region.

3.2.1.3 Variations due to positioning of the meter

Errors in kV measurements may also be due to the positioning and orientation of the meter. Figure 3.2.1.3.1 shows four measurements made at different distances away from the focal spot, i.e. the meter was moved up and down in the beam. The graph shows that this makes little difference to the kV measured, the standard deviation of the four measurements was 0.05 kV at a nominal tube voltage of 28 kV which was measured to be 29.2 kV. This gives a coefficient of variation of 0.2%. Except in cases where the output of the tube is particularly low, and the meter needs to be moved closer to the focal spot to make sure that the radiation intensity is high enough for a measurement to be made, the meter is always the same distance away from the focal spot since it usually sits on the breast table during measurements, so this tends not to be a source of error.

Figure 3.2.1.3.1
Variation in measured kV with distance of meter from the focal spot



spot, the meter position can vary across the breast table both parallel to the anode-cathode axis and perpendicular to it, and the meter can be rotated so that it is oriented with the display pointing in different directions. The geometry of the measurements of the position of the sensor with respect to the X-ray field are shown in figure 3.2.1.3.2

Figure 3.2.1.3.2 - The orientation of the axes

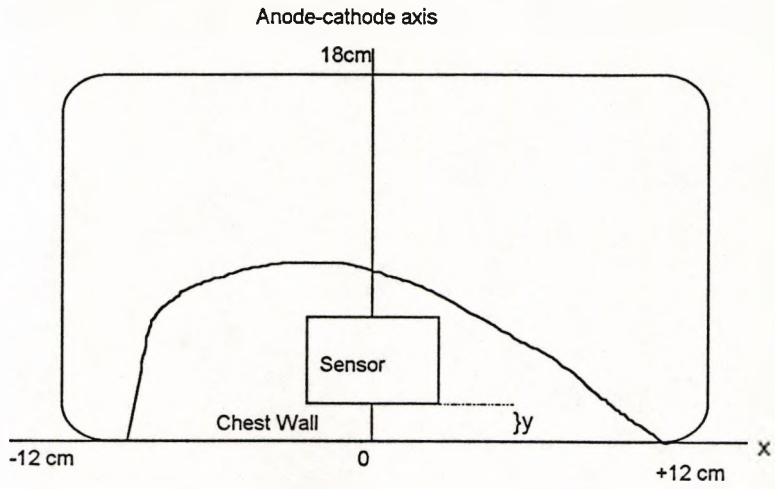


Figure 3.2.1.3.3 shows how the measured kV varies as a function of the position of the meter in a direction perpendicular to the anode-cathode axis. The measurement was made twice at each position and the average used for the graph. A negative distance denotes the sensitive area of the meter positioned to the left of the centre line and a positive distance to the right. These measurements give a coefficient of variation of 0.2%.

Figure 3.2.1.3.3
Variation of kV measured perpendicular to the anode-cathode axis, zero is central

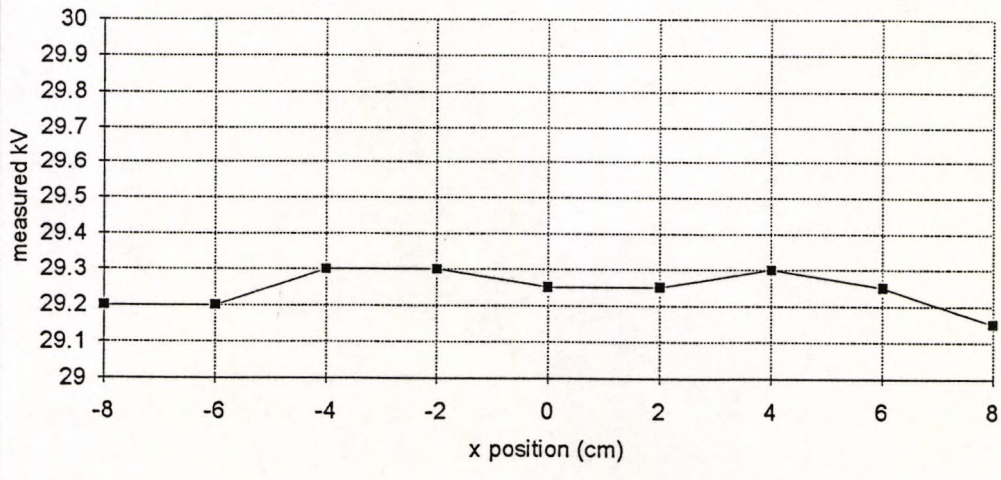
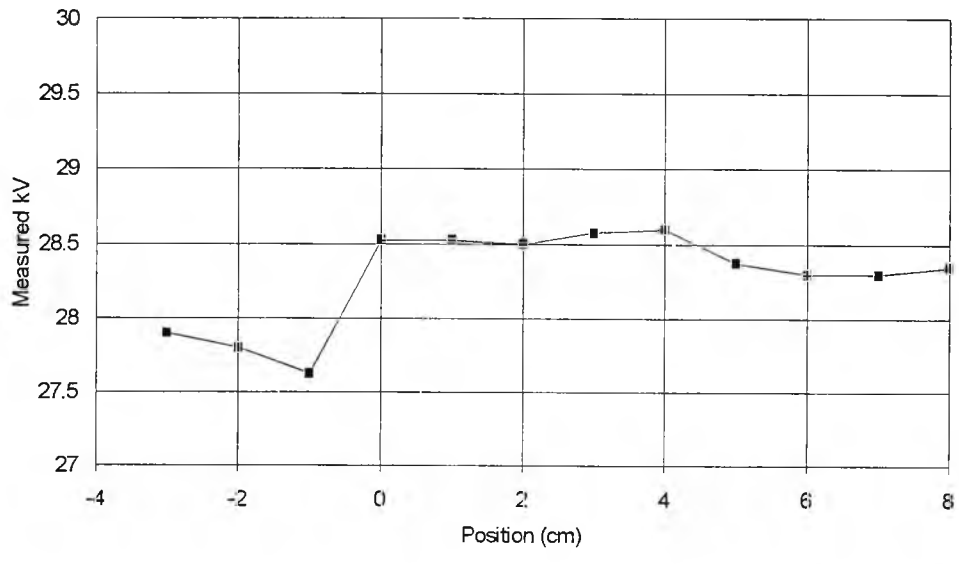


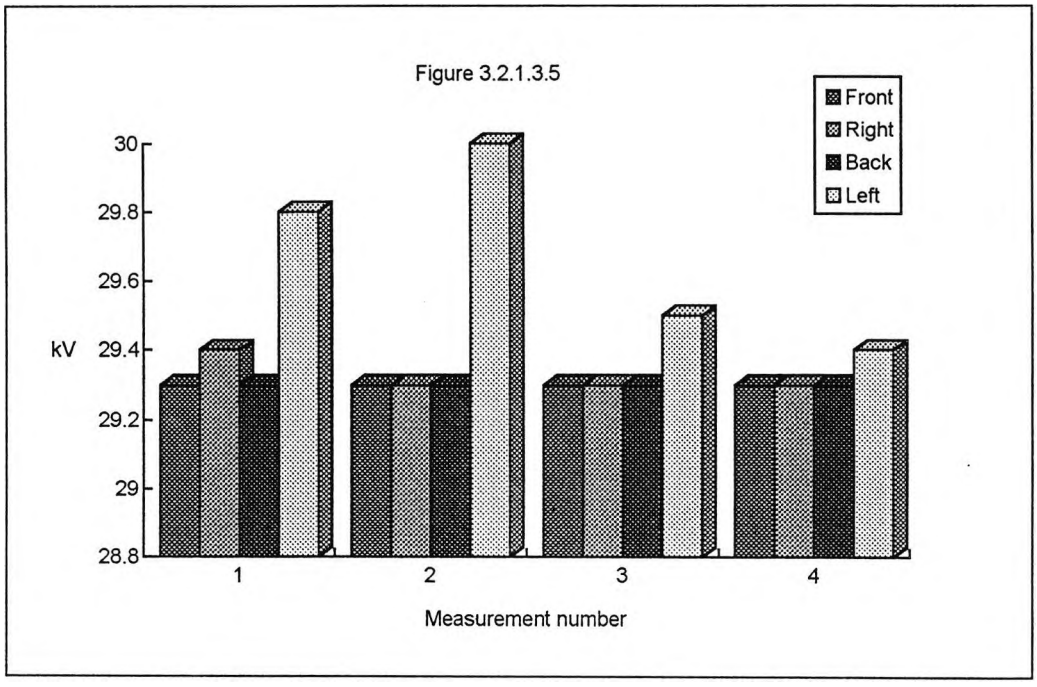
Figure 3.2.1.3.4 shows how the meter reading varies along the anode cathode axis. The chest wall edge of the beam is the zero of this axis with the positive direction going into the beam. The results indicate that the positioning of the meter is not critical so long as the sensor lies within the area the beam. In routine measurements, the sensor was positioned 2 cm away from the chest wall edge. The estimate of kV variation will therefore be based on values of x greater than or equal to zero to reflect the true measurement situation, and is equal to 0.4%.

Figure 3.2.1.3.4

kV variation with position along anode-cathode axis



The effect of the orientation of the meter on the measured kV was also tested. The meter was placed on the breast table with the sensitive area central to the beam with the edge of the sensitive area 2 cm in from the front edge of the beam. The kV was measured four times in each position and the results are shown in figure 3.2.1.3.5. There is very little variation in the measured kV except when the meter is facing to the left. It is not clear whether these variations are due to random fluctuation in the voltage across the X-ray tube or whether they reflect the meter function. Consequently, this experiment was repeated on a different X-ray machine with the same meter, ten measurements were made for each orientation. All forty measurements were identical which seems to indicate that the variations were not due to the meter, however, the data used to generate figure 3.2.1.3.5 gives a coefficient of variation of 0.7% which will be used to provide a worst case estimate of the accuracy of the kV measurements.



3.2.1.4 Estimate of error in kV measurements

The theory of propagation of errors [111] allows errors from all sources to be combined to give an overall estimate of the error in the final result. When quantities are multiplied or divided to give a final result, the expression below gives the relationship between the precision (standard deviation) in the final result and the precision of all the contribution components.

$$\left(\frac{\delta kV}{kV}\right)^2 = \left(\frac{\delta a}{a}\right)^2 + \left(\frac{\delta b}{b}\right)^2 + \ldots\ldots \qquad \text{Equation 3.1}$$

Where δa is the estimated error due to a particular source and therefore $\frac{\delta a}{a}$ is the coefficient of variation (not expressed as a percentage) for that particular source of error. For the sources of error which have been identified in this chapter, this gives a result of

$$\left(\frac{\delta \mathrm{kV}}{\mathrm{kV}}\right)^2 = 0.003^2 + 0.002^2 + 0.002^2 + 0.004^2 + 0.007^2$$

$$\left(\frac{\delta \mathrm{kV}}{\mathrm{kV}}\right)^2 = 0.000082 \qquad \text{Equation 3.2}$$

$$\left(\frac{\delta \mathrm{kV}}{\mathrm{kV}}\right) = 0.0091$$

or 0.9% coefficient of variation overall for kV measurements.

3.2.2 Output

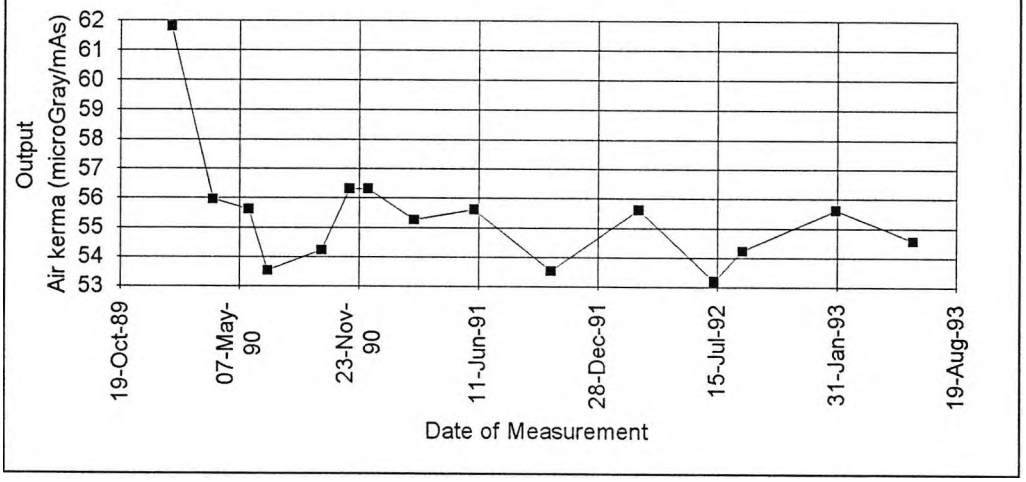
In the same way as the error in kV measurements was evaluated for the three different sources, this procedure was also followed for output measurements.

3.2.2.1 Variation due to statistical fluctuations of the X-ray unit

The magnitude of output fluctuations were measured by recording the time series of several X-ray units. One which shows no long-term drift was chosen and the range and standard deviation of the measurements calculated. This is shown in figure 3.2.2.1, the coefficient of variation calculated from this is 3.7% using all the available data, or 1.9% if the first point, which appears to be untypical, is ignored. The initial fall in output has been observed in other units both within this study and outside of it [135] and appears to be normal .

Figure 3.2.2.1

Central and East London Mobile 1 - Ouput at 28 kV



3.2.2.2 Variations due to the dosemeter

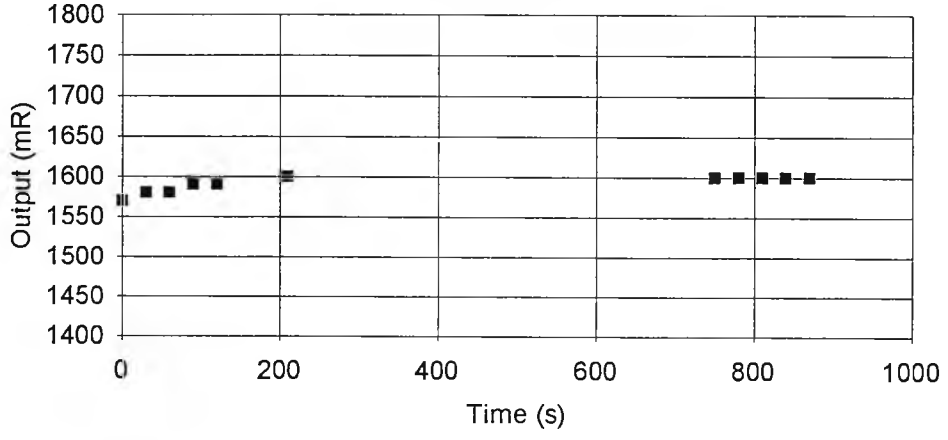
Variations in output measurements due to the performance of the dose meter were categorised in the following way, and the error due to each category was measured:

a) instrument warm up,
b) ion chamber saturation effect,
c) random fluctuations (set of 5 measurements).

The effect of instrument warm up is illustrated in figure 3.2.2.2 which

shows a sequence of five measurements, all of 28 kV and 40 mAs, made at thirty second intervals, another made ninety seconds later and after a gap of 10 minutes a further five measurements also at thirty second intervals (during the gaps the X-ray tube was making other exposures which were being measured with the dosemeter). This illustrates that the meter reaches a plateau between the fifth and eighth exposure and that the maximum error which can be expected due to insufficient warm up time is 1.9%.

Figure 3.2.2.2
The warm up of a dosemeter



Ion chamber saturation only occurs at very high dose rates. Because of the low kV used and the modest tube currents which the X-ray tubes are capable of producing, there is no possibility that the ion chamber would become saturated.

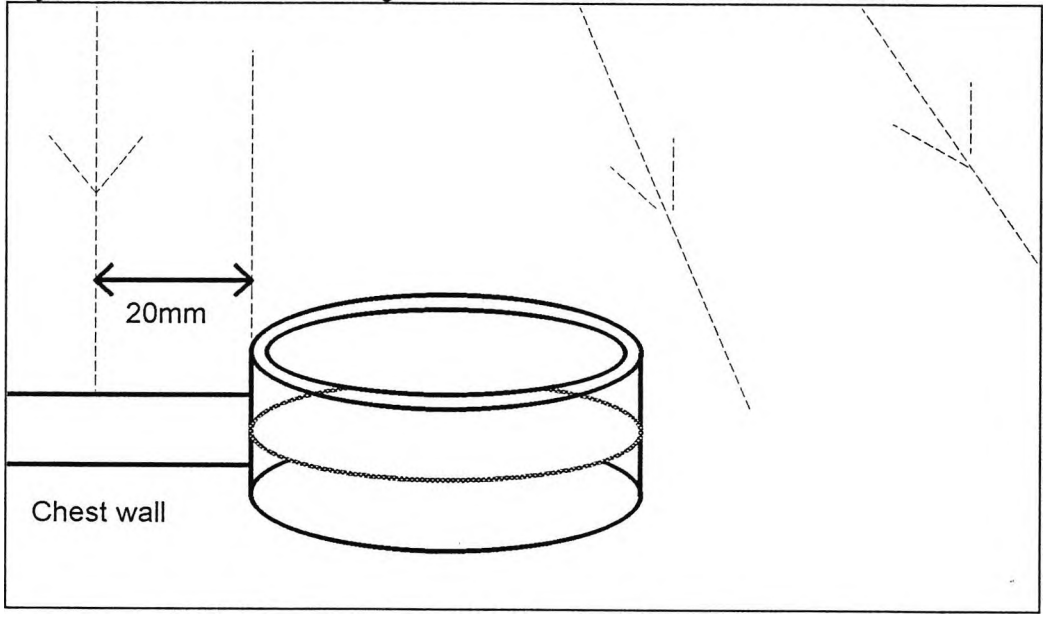
For a set of 5 repeated measurements, the coefficient of variation is usually 0% and depends on the machine being tested, the worst case for a set of five measurements has been chosen to give a conservative estimate of the accuracy of output measurements of 0.4%. This shows that any variations in the response of the meter over a short period of time are negligible as long as the meter has been stabilised by an initial exposure before the measurements begin and has been given enough time for the ion chamber to warm up to room temperature.

3.2.2.3 Variations due to experimental technique

The experimental technique can also alter the results of output measurements, Positioning (x, y and z) has a very strong influence on

the measured output. The position of the dosimeter ion chamber is set so that when the chamber is resting on the breast table, the edge of the chamber from which the support arm emerges is 20 mm inside the light beam, as shown in figure 3.2.2.3.1. The ion chamber is then left stationary on its supporting tripod and the X-ray unit lowered so that the red line on the edge of the ion chamber is 400 mm away from the marked position of the focal spot. The definitions of the axes used will be that x is perpendicular to the anode-cathode axis in the horizontal plane, y is parallel to the anode cathode axis and z is vertical. The x positioning is done by eye setting the ion chamber over the centre of the field marks, this can be reproduced to within 5 mm. The y positioning is done using a ruler and can be done to better than 3mm (errors occur mainly due to the unsharpness of the light beam), the z positioning is done with a tape measure and is subject to parallax errors in the positioning of no greater than 5 mm.

Figure 3.2.2.3.1 Positioning of the ion chamber within the beam



In order to evaluate how much the positioning of the ion chamber contributes to the overall error in output, the position of the ion chamber was altered in a controlled manner and the variation in the output measurement was plotted on a graph and the best fit equation found. The z dependence (the distance between the marked focal spot and the reference line on the ion chamber) theoretically obeys an inverse square law relationship, the actual relationship will also be subject to any differences between the marked focal spot position and the true focal spot position. The uncorrected output was plotted as the dependent

variable and $1/(a+z)^2$ was plotted as the independent variable, this appears in figure 3.2.2.3.2. This form of equation allows the parameter representing the offset of the focal spot from its marked position (a) to vary as well as the parameter representing the tube output (k), this equation

$$Output(x,y,z) = \frac{k}{[a+z]^2} \qquad \text{Equation 3.3}$$

is then equivalent to a straight line of the form

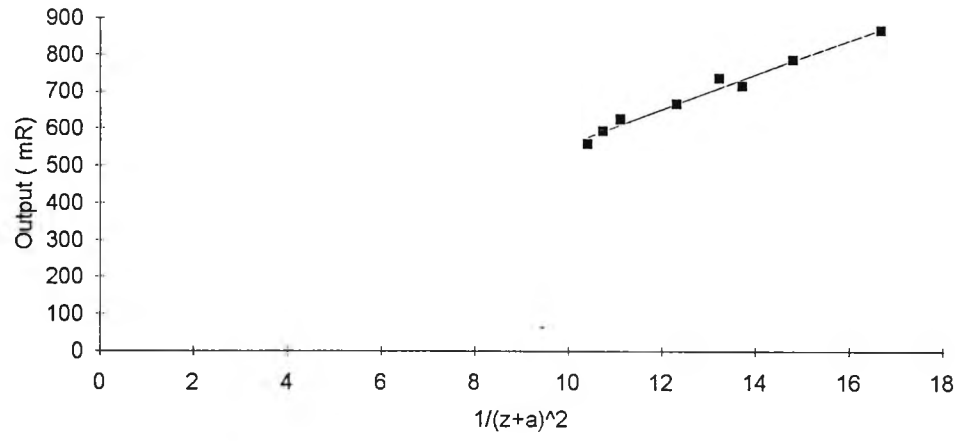
$$y = mx + c \qquad \text{Equation 3.4}$$

This was fitted to the data using a least squares fit giving an equation of best fit

$$Output(x,y,z) = \frac{72.04}{(z+0.043)^2} \qquad \text{Equation 3.5}$$

Figure 3.2.2.3.2

Inverse Square Law Variation of Output - Data (points) and theoretical (straight line)



which means that for a standard position of 400 mm focus to chamber, a positioning error of +5 mm gives an output error of -2.1% and -5 mm gives an output error of +8.7%.

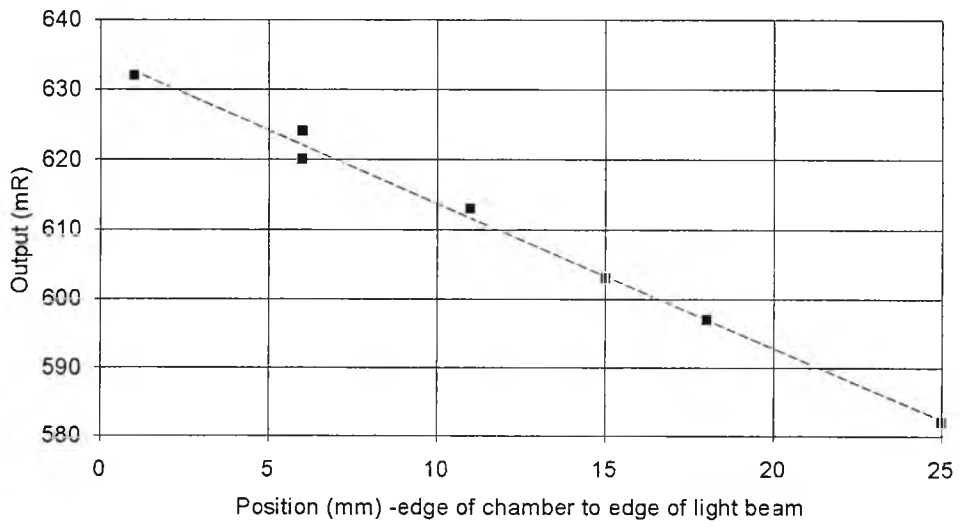
The y dependence was also estimated by measuring the output while moving the position of the ion chamber along the y axis. One expects there to be a non- uniformity of the field due to geometrical factors and also to the anode heel effect such that the intensity falls off the further away the chamber is moved from the chest wall. There will also be a dramatic fall in the field intensity as the beam edge is encountered. The position of 20 mm in from the chest wall edge was chosen to avoid this scenario even if the light and X-ray beams were poorly aligned. The ion

chamber is 60 mm in diameter and therefore, the response will not follow the beam profile exactly but will show the radiation profile averaged over the area of the ion chamber. This makes lateral positioning errors relatively unimportant in the overall measured output. Fitting the data to a straight line shown in figure 3.2.2.3.3, the output dependence is given by

$$Output = 635 - 2.09y \quad (y \text{ is in } mm) \qquad \text{Equation 3.6}$$

So assuming that an error in the y position of 3 mm has been made, the error in the output is 6 mR in 593 mR which is 1%.

Figure 3.2.2.3.3
Dosemeter Measurement vs position along a/c axis



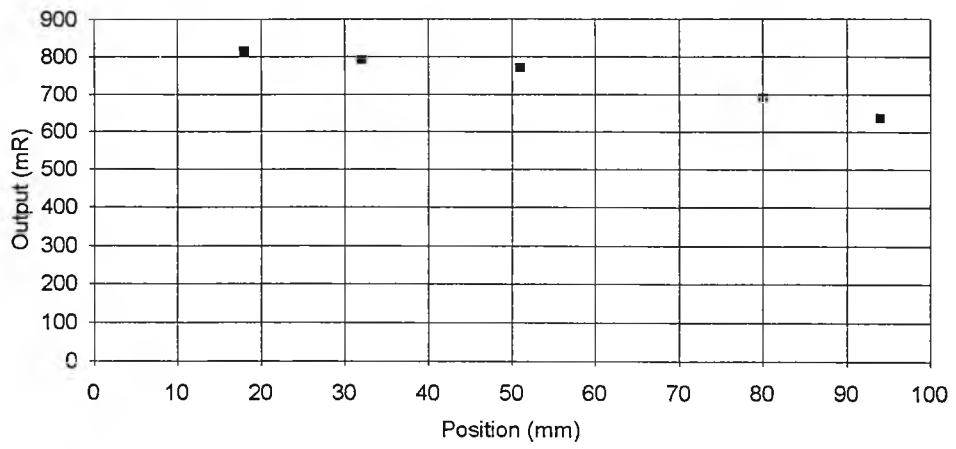
Similarly in the x direction i.e. perpendicular to the anode cathode axis, a small variation due to geometrical factors is expected. The output as a function of the x position is shown in figure 3.2.2.3.4. Using the inverse square law, the relationship between exposure and position is expected to be

$$Output = Output_{centre} \times \left(\frac{FFD^2}{FFD^2 + x^2} \right) \qquad \text{Equation 3.7}$$

The best fit was for an FFD of 177mm and with an output of 825 mR at x=0. From this, if there is a positioning error of no more than 5mm in the x direction, the error in the measured output will be 0.1% or less.

Figure 3.2.2.3.4

Variation in dosemeter response perpendicular to the anode-cathode axis

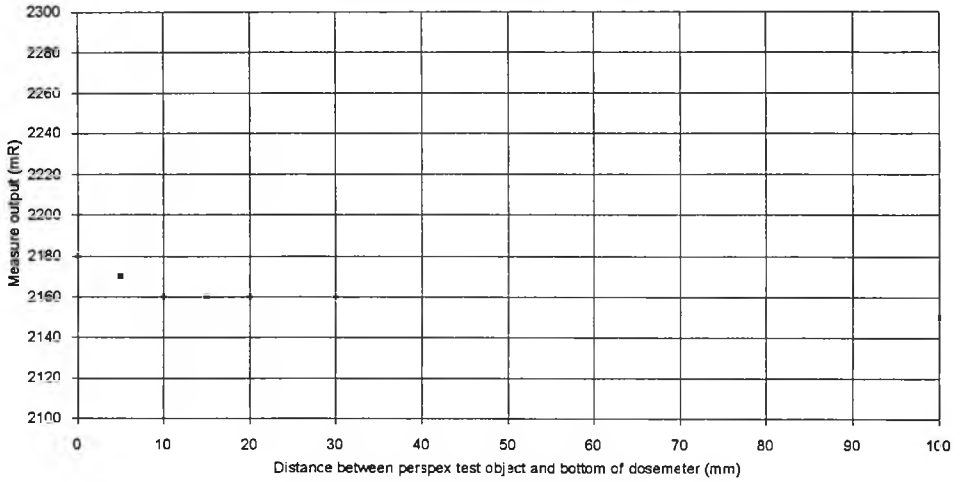


The output of the tube ought to be linearly dependant on the mAs, this is not always the case and is a quantity which is routinely measured. This can be due to two reasons, the voltage wave form (dependent on the rectification of the unit), and therefore the energy spectrum of the X-rays, may fluctuate throughout the exposure giving a variation which is due to non-linearity of the dosemeter with photon energy or it may be due to switching errors either in the control circuitry of the X-ray unit or of the dosemeter itself. These effects cannot be separated but the magnitude of these effects combined together was evaluated from a typical X-ray unit which gave a coefficient of variation of 3.7%.

To evaluate the effect of backscatter on air kerma measurements, the position of the ion chamber in an X-ray field was kept constant and blocks of perspex were placed underneath it to act as a scattering medium. The graph in figure 3.2.2.3.5 shows that the effect of backscatter is small beyond a distance of 10 mm between the scattering medium and the bottom of the ion chamber, it also shows that by 75 mm separation, the measured air kerma has reached its plateau. Consequently, if the ion chamber is more than 75 mm above the breast table, the measurement will be free of scattered radiation at 30 kV. If there is no gap and full backscatter is incorporated in the measurement, the increase in measured value due to backscatter is 30/2150 = 1.4% which represents the maximum possible error due to backscatter.

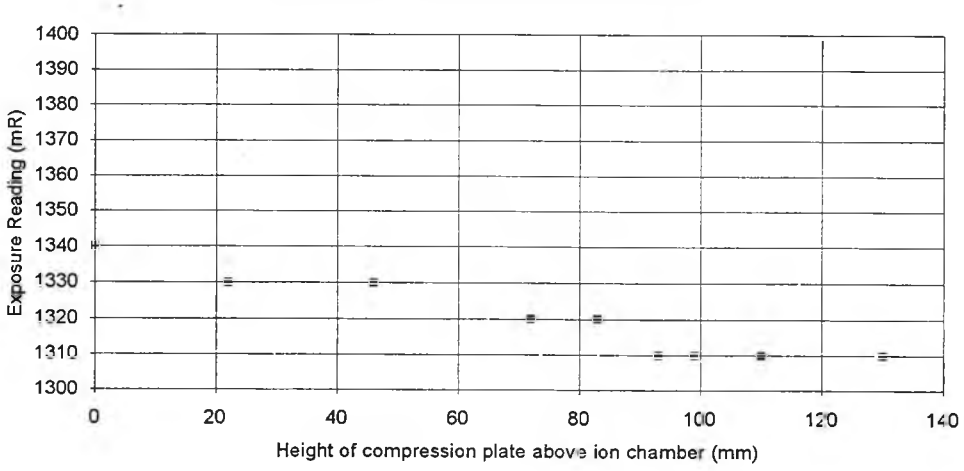
Figure 3.2.2.3.5

The effect of backscatter on output measurements



There will also be some forward scattered radiation from the compression plate which will influence the exposure recorded by the ion chamber. In order to evaluate this, the height of the ion chamber and position in the beam was kept constant and the compression plate was moved up and down the z axis. The graph in figure 3.2.2.3.6 shows how the height of the compression plate above the ion chamber affects the results. The range is 30 in a measurement of 1320 = 2.3% giving a typical error of ± 1.2%.

Figure 3.2.2.3.6 The effect of forward scattered radiation from the compression plate on air kerma measurements



3.2.2.4 Correcting/normalising for kV

If a time series of output is plotted, it is always assumed that the kV is constant, this is true in the majority of cases, but if there is a drift in tube kV or if the kV is reset by the engineer for any reason, the results are not

strictly comparable. One can either assume that the theoretical relationship that output is proportional to kV^2 and make corrections accordingly or one can perform a curve fitting to the data. The theoretical relationship will be in slight error for any system which incorporates filtration, because the spectrum arriving at the detector is altered by the filtration but is a good enough approximation for most purposes. This also allows all output measurements to be normalised to exactly 28 kV making inter comparisons of different X-ray units valid. The resulting figure, achieved by taking the output per mAs at one metre and dividing by the square of the tube voltage, has been called the *output coefficient*.

3.2.2.5 Correcting for filtration

Similarly, there may be a need to allow for changes in filtration which may occur over time. Increased filtration reduces the output and produces a harder beam. Half value layer (HVL) measurements are only accurate to within 0.01 mm in 0.3 mm (3.3%) at best and the changes measured have been smaller than this and could be explained by statistical variation alone, therefore no corrections were made for changes in HVL.

3.2.3 Automatic Exposure Control

The automatic exposure control device serves the function of ensuring that the density of the final image is constant regardless of the settings of any of the exposure parameters, such as kV selected or whether or not the grid is being used, and irrespective of the anatomical variations of the patients.

3.2.3.1 Random Fluctuations

These results are based on radiographers daily measurements of the AEC function. Data from Southend Basildon and Thurrock since the tube was replaced gives a straight line with time with a gradient of about 10^{-5}, which essentially shows no drift. This data has been used to derive statistics for the random variation of both mAs and optical density, this is shown in table 3.2.3.1. The coefficient of variation for mAs is 3.4% and for optical density is 6.9%. The variation of mAs represents the random variation of the AEC for a stable system with attenuation, geometry and X-ray spectrum held constant; the variation of OD incorporates the

variations in processing too, one would therefore expect the coefficient of variation in the processing system to be 6.0%, this is measured independently by light sensitometry.

	mAs	O.D.
mean	32.39	1.393
standard deviation	1.16	0.096
coefficient of variation	3.6%	6.9%

Table 3.2.3.1 Results from daily measurements at Southend Basildon and Thurrock following tube replacement

3.2.3.2 Variations due to the functioning of measuring equipment.

When the AEC is tested, the measuring equipment is the AEC itself and therefore this is identical to the quantity measured in section 3.2.3.1. so the variations of the X-ray unit are in fact the same as the variations of test equipment shown in table 3.2.3.1 previously.

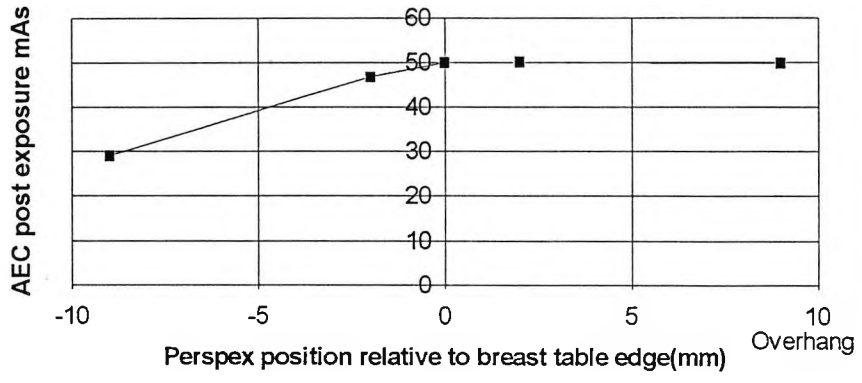
3.2.3.3 Variations due to measurement technique

The position of the perspex block may influence the final result due to the amount of scatter reaching the sensor. A test was done moving the edge of the block of perspex along the anode-cathode axis so that it extended beyond the chest wall edge by varying amounts and making exposures under automatic exposure control at each of those positions. These results are shown in figure 3.2.3.3 and indicate that as long as the perspex block is level with the edge of breast table or overhangs it by a positive amount, the position of the perspex is not critical.

The presence or absence of a compression plate may also influence the function of the AEC because it alters the spectrum somewhat by removing the low energy end of the radiation spectrum. In order to eliminate this source of variation, the compression plate was used for all measurements of AEC function. Likewise, the grid was in position for all measurements except those specifically designed to measure the effect of the grid, and the detector was always in the position closest to the chest wall.

Figure 3.2.3.3

The effect of perspex position on AEC test



The thickness and density of the perspex test blocks can vary due to variations in the manufacturing process. For the physics measurements, the same set of perspex plates taken from the Leeds test object was used throughout. The radiographers tests were done on a local set of perspex blocks which again were always the same for that particular unit. This means that results from one unit to another are only comparable for measurements made by the physicist but are internally consistent. All sets of data are kept separate so there are no errors from this source except where data is drawn from more than one set of measurements i.e. there may be slight errors when comparisons are made.

The AEC controls the density on the film, consequently, the important parameter is not the mAs, but the optical density at a fixed position. This is influenced in addition by the condition of the processor, the cassette and the film. A cassette was chosen at each site and designated as the QC cassette, so again, only inter-comparisons between units present a problem. In order to allow for changes which occur in density because of either the processor or the film itself, a sensitometric film was produced for each set of measurements which was used when necessary to correct for the combined effect of film and processor on the final density. The effect of the two components on the film cannot be separated out.

3.2.4 Half Value Layer

The half value layer gives an indication of the energy spectrum of the X-ray beam. The higher the proportion of short wavelengths in the X-ray beam, the greater will be the half value layer, because high energy, short wavelength

photons are able to penetrate materials more easily, therefore more material is required to reduce the intensity of the beam by a factor of two. The half value layer is influenced by changes in kV and also by changes in filtration. A variation in HVL demands further investigation in order to ascertain the cause of the change.

3.2.4.1 Random fluctuations

HVL - no drift is expected, theoretically the HVL can only change if the filter is damaged therefore any variations observed can be ascribed to limitations of the measurement technique. For a series of eleven HVL measurements on one unit, the coefficient of variation was measured to be 4.4%.

3.2.4.2 Equipment Variations

The analysis for this section is exactly the same as for the exposure measurements because the same dosemeter was used for both measurements. When performing HVL measurements, the zero reading is repeated at the end of the sequence to check that there has been no drift. The measurements are made after the tube output has been measured therefore the dosemeter should have warmed up and reached a stable value by the time these measurements were made.

3.2.4.3 Variations due to measurement technique

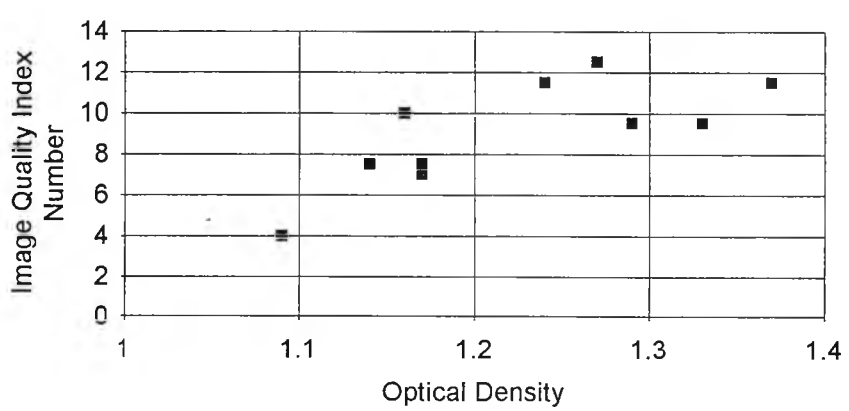
One source of variation due to measuring technique is the thickness of the filter. It is necessary to determine the thickness of each individual Al filter, in order to determine what difference the sequence makes. It was found that one of the two 0.5mm filters was not the correct thickness, this filter was taken out of use. The other filters were sufficiently accurate that no measurable difference was observed regardless of the filter used. Collimation is used to reduce scatter and should be consistent for all measurements. The correct positioning of filters is also important because of the effect of scatter reaching the detector. One particularly important variable is the position of the detector because the anode heel effect gives varying radiation quality with position, however, the position of the detector is not changed following the output measurements so the error due to positioning will be the same as that for output measurements.

3.2.5 Image Quality

3.2.5.1 Random Fluctuations in Image Quality

In order to evaluate this, a time series where there is no overall drift is required. The size of the fluctuations may be large and will be determined by many effects as well as the true functioning of the equipment. It is impossible to separate out true variations from variations caused by the scorer or the set-up of the equipment, taking data from the Central and East London Mobile, which has shown itself to be very stable as far as physical parameters are concerned, the overall variation in image quality score is shown in figure 3.2.5.1 as a function of the optical density of the film. There is a strong correlation between image quality and optical density indicating that in an otherwise stable system, the AEC function and the film processing are the most important factors determining image quality.

Figure 3.2.5.1 Relationship between image quality and optical density



3.2.5.2. Variability of scoring of the Leeds TOR(MAX) test object

Test objects are the most common way of assessing image quality. They are semi-subjective but are considerably easier to administer than objective techniques such as ROC analysis [136]. The main questions regarding their use are firstly, can the absolute scores be used with any confidence or is their value restricted to comparisons, this depends on the variability of scoring.

Method

In order to determine the coefficient of variation of scoring the Leeds TOR(MAX) test object, the following test was done. Ten exposures were made of the Leeds TOR(MAX) test object under manual control. The exposures were all at 28 kV, 80 mAs nominal on a Siemens Mammomat 2S which had been shown to a have a coefficient of variation of 0% in kV. The films were exposed with a 30 second break between exposures to allow the tube to cool, in order to reduce the risk of thermal damage to the anode. There are no physical reasons to suspect that the image quality would change as a result of temperature changes of the anode, but the safety procedures which had been followed co-incidentally ensured that this would not be a problem. They were then processed as part of a batch in order to minimise processor variations from film to film. To check this, the optical density of the tenth step on the wedge was measured and was constant within the measurement limits of the densitometer.

The films were allocated numbers one to ten in the order in which they were produced. In order to ensure that on the following days, they were re-scored in a random sequence, they were allocated numbers from a random number table. A person who was not involved in the scoring renumbered the films each day and the scorer was not allowed to refer to film sequences or scores on previous days.

The images were scored by a person who routinely uses this particular test object and could be said to be an experienced scorer. Each item within the test object, 107 in all, was scored on the 0, 1/2, 1 basis; an object clearly seen scores 1, any object which is not seen scores 0 and if the scorer was unsure, a score of 1/2 was assigned.

Nine items were given a score: the 6 mm circles (0-12), the 0.5 mm circles (0-9), the 0.25 mm circles (0-9), the high contrast resolution patterns parallel and perpendicular to the anode cathode axis (0-26 each), the low contrast resolution pattern (0-10) and the large, medium and small micro calcifications (0-5 each).

Results

The tables 3.2.5.2.1 and 3.2.5.2.2 show the total score for each film on each day, figure 3.2.5.2.1 includes the scores for the calcifications which are recommended only as a subjective measure whereas table 3.2.5.2.2 does not, hence the lower scores (scoring out of a maximum of 92 instead

of 107). The mean and standard deviation by film are in the right hand columns of the tables and the mean and standard deviation by day are on the bottom rows. The overall mean and standard deviation are in the bottom right hand corner. When the calcifications are included, the coefficient of variation is 2.43/78.23 =3.1% and when the calcifications are excluded, the coefficient of variation is 1.73/67.39 = 2.6%. Including the calcifications in the score reduces the reproducibility of the measurements. A measurement error of +/- 2 in IQIN (one standard deviation) can be expected for any one film.

Film	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10	mean	sd
1	79.5	76	76.5	80.5	75.5	83	79.5	81.5	82.5	81.5	79.6	2.74
2	77	78	77	80.5	81	77	78.5	80.5	81	78.5	78.9	1.70
3	74.5	76	75.5	76	78	77.5	78.5	75.5	79.5	78	76.9	1.61
4	76	75.5	77.5	77	76.5	75.5	78	78.5	80.5	80	77.5	1.76
5	75	75.5	72.5	76	77	74.5	77.5	78	79	80.5	76.55	2.34
6	74	77.5	77	77.5	78	77.5	78	77.5	78.5	80.5	77.6	1.60
7	73.5	76.5	75	75	80	79.5	77.5	81	80	81.5	77.95	2.83
8	75.5	75.5	78.5	81.5	80	81	78	80	81.5	85	79.65	2.91
9	76.5	78	79	78.5	81.5	80	81.5	82	81.5	82.5	80.1	2.01
10	75.5	75.5	77.5	75	78.5	76	79	78	78	82.5	77.55	2.24
mean	75.7	76.4	76.6	77.75	78.6	78.15	78.6	79.25	80.2	81.05	78.23	
sd	1.72	1.05	1.88	2.40	1.98	2.67	1.20	2.07	1.46	2.05		2.43

Table 3.2.5.2.1 Summary of image quality scores with calcifications included

Film	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10	mean	sd
1	68.5	67	67.5	67.5	66.5	72	69.5	69.5	70.5	70.5	68.9	1.79
2	65	67	68	69.5	68	68	67.5	68.5	69	67.5	67.8	1.23
3	65.5	67	66.5	67	66	65.5	66.5	66.5	67.5	66	66.4	0.66
4	65	66.5	65.5	65	66.5	65.5	66	66.5	68.5	68	66.3	1.18
5	66	65.5	63.5	66	66	64.5	66.5	67	67	68.5	66.05	1.38
6	65	66.5	66	66.5	67	67.5	68	66.5	67.5	67.5	66.8	0.89
7	64.5	65.5	65	66	68	67.5	65.5	68	67	69.5	66.65	1.60
8	67	65.5	68.5	70.5	70	69	68	69	69.5	73	69	2.03
9	67.5	67	68	68.5	70.5	69	69.5	71	68.5	69.5	68.9	1.26
10	66.5	65.5	67.5	66	66.5	67	67	67	67	70.5	67.05	1.34
mean	66.05	66.30	66.60	67.25	67.50	67.55	67.40	67.95	68.20	69.05	67.39	
sd	1.30	0.71	1.60	1.75	1.62	2.17	1.37	1.54	1.21	1.99		1.73

Table 3.2.5.2.2 Summary of image quality scores with calcifications excluded

Figures 3.2.5.2.1 and 3.2.5.2.2 show how the mean score for 10 films varies from day to day (with and without the calcifications included in the score). In both cases, there is an upward trend in the score with time. This

must be a learning effect of some kind which is surprising because the scorer is accustomed to the test object. Figure 3.2.5.2.3 shows the same learning effect with calcifications, which would be expected since these are not normally scored, and indeed the scorer was aware of the process as criteria for detection were being established; however, even when the calcifications have been subtracted from the total, a learning effect is still evident. The implication of these results is that even for an experienced scorer, in excess of 100 films need to be scored in order to reach the plateau of the learning curve.

Figure 3.2.5.2.1 TOR(MAX) mean score for all ten films by day

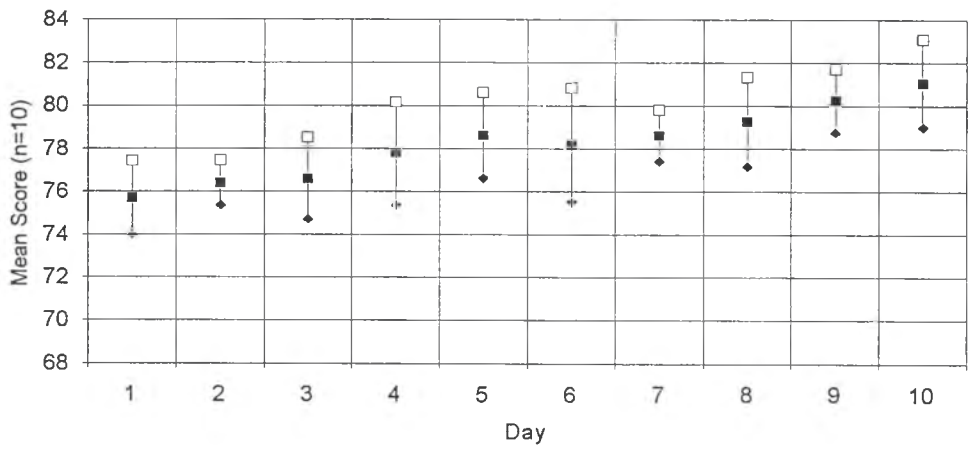


Figure 3.2.5.2.2 TOR(MAX)mean score excluding calcifications for all ten films by day

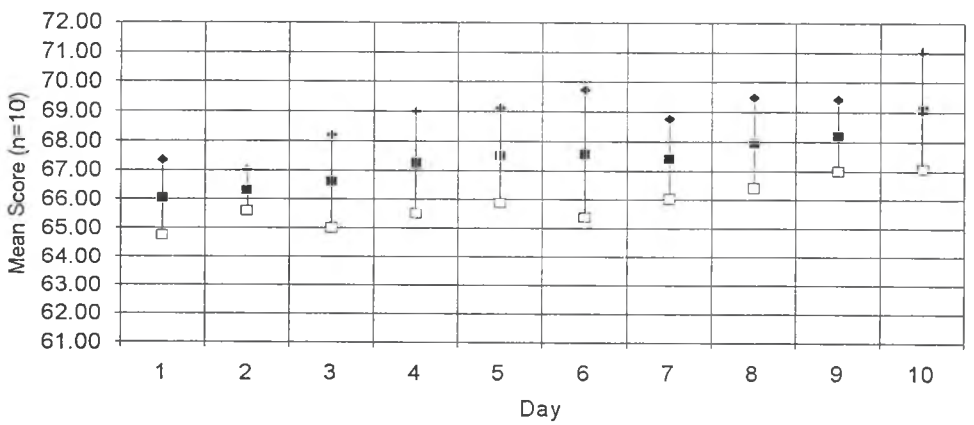


Figure 3.2.5.2.3
Small Calcifications

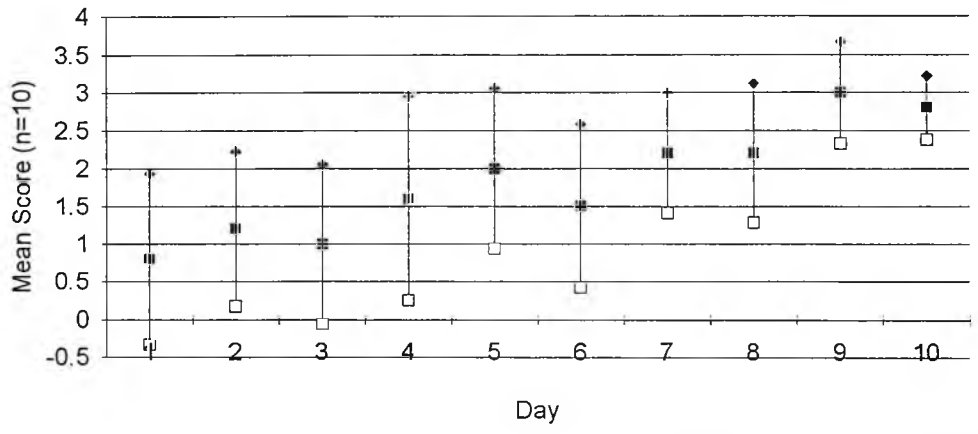


Figure 3.2.5.2.4 shows the mean score for ten films (excluding calcifications) averaged over the ten days of scoring. The error bars from film to film would suggest that some of the variations (e.g. between film 3 and 5) may not be explained by random variations and could reflect a genuine difference in the quality of the images. Figure 3.2.5.2.5 is a plot of the mean score against the optical density of one step on the step wedge (each point measured four times). There is no correlation between the differences in image quality score and the optical density. This is to be expected because the optical density variations were negligible. If the differences are real, they cannot be ascribed to changing optical density. Such changes could possibly have occurred if the kV was fluctuating slightly, but previous measurements showed the kV to be very consistent (C.V. = 0% for 5 measurements) and this is unlikely to be the case. It would appear that the variations which are seen are random variations. One possible cause is poor film screen contact, little time was allowed between loading the cassette and making the exposure, in clinical practice there would be approximately four minutes between one woman and the next allowing time for any trapped air to leak out.

It was concluded that the variability in re-scoring a film was so great that only gross changes could be identified with any confidence. Because of the learning effect which was observed with an experienced scorer, it was also concluded that comparisons scored at the same time may be considered to be valid but that otherwise care needs to be taken to account for the learning effect.

Figure 3.2.5.2.4 Mean total scores from TOR(MAX) excluding calcifications

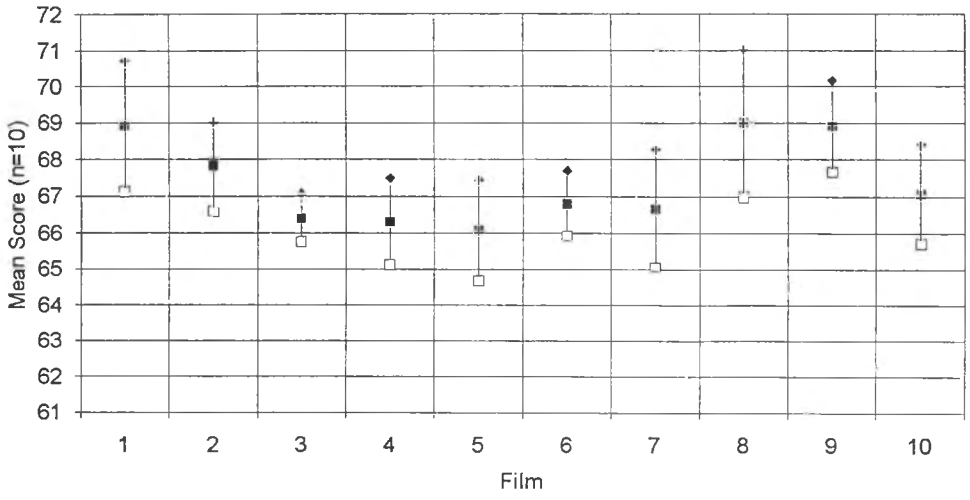
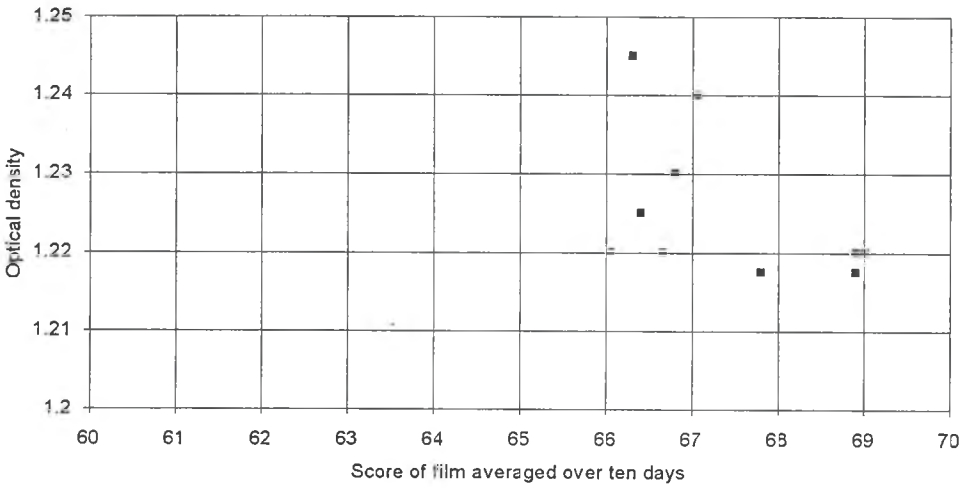


Figure 3.2.5.2.5

Variation in image quality score with optical density



3.2.5.3 Image Quality variations due to changes in optical density

In order to make a valid comparison between the same phantom used on different units, it is necessary to make corrections for changes in the default optical density, and for kV variations at a nominal 28 kV. Images have been produced on one unit at a range of optical densities and in a separate experiment on a different unit at a range of kV settings in order to evaluate the dependence of the score on the optical density and kV. These results are shown in figures 3.2.5.3.1 and 3.2.5.3.2.

Figure 3.2.5.3.1

Image Quality score at 28 kV

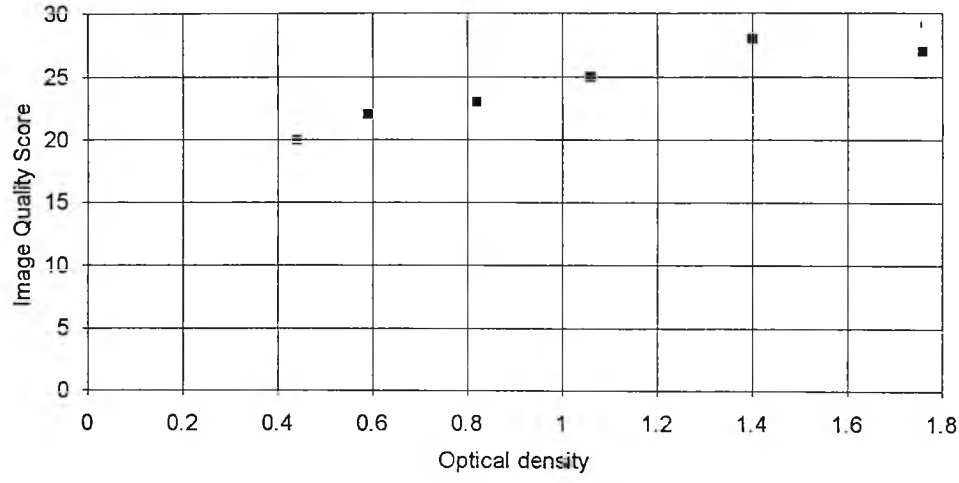
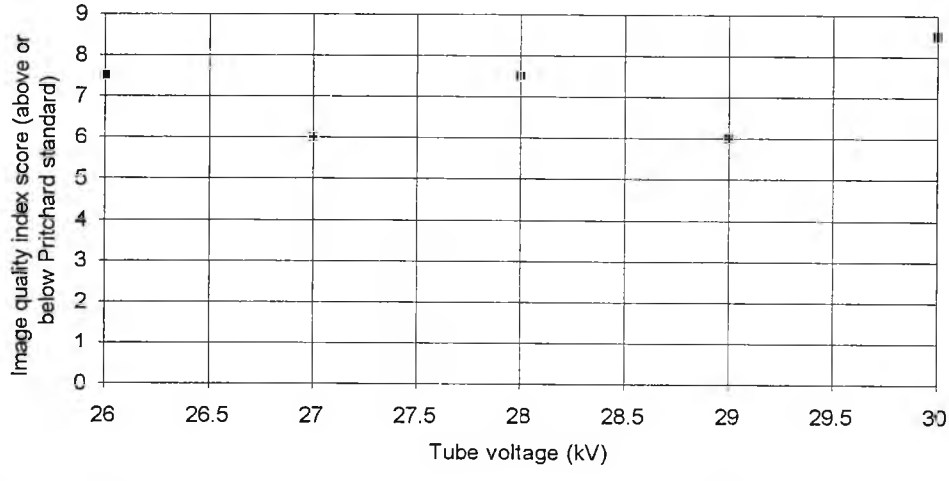


Figure 3.2.5.3.2 Variation of image quality with kV set



There is no significant change in image quality with tube voltage in the diagnostic mammography range but an improvement with increasing optical density is seen.

3.2.6 Focal Spot Size

3.2.6.1 Focal Spot Dimensions

Taking a series of seven measurements on one mobile X-ray unit, the coefficient of variation was 20.6% for the length and 13.5% for the width. These measurements incorporate errors from the measurement technique too, but the difference in the CV for length and width is probably due to

the difficulty of accurately positioning the jig at such an angle that it was properly aligned with the central axis. The measured length depends on this but the width does not, therefore 13.5 % is a reasonable estimate of random error for both focal spot length and width.

3.2.6.2 Changes in measured focal spot dimensions due to equipment

The equipment used, a jig to support the star and the slit, and the film, contain no electrical parts or mechanical parts which are involved in the measurement, any errors are due to the way in which the equipment was used rather than the equipment itself.

3.2.6.3 Errors in the focal spot size due to measurement technique.

The magnification which is used in order to be able to measure the focal spot size depends upon the height of the supporting jig. The actual value of magnification was measured on the test film in comparison to known dimension on the star. This value is part of the calculation so although magnification may change, this is automatically compensated for. Lateral misalignment of a few millimetres is possible, this should not cause any error in the length measurement but may affect the width, this variation is included in the 13.5 % estimated for random variation. The angle at which the measurements were made will have negligible effect on the width measurement but a very marked one on the length. Although this cannot be measured directly, it can be inferred from the measured coefficient of variation as 15.6%. Sometimes the angle of measurement is incorrectly set, but the size of the error is known in this event corrections can be made for the measurement angle.

3.2.7 Alignment

3.2.7.1 Random variations in alignment

In general, there should not be any random variations in the equipment, changes must be due to special causes. If the system has been re-aligned by the engineers or jolted loose in transit (which is unlikely because once in position, the tube is tightly clamped down) changes may be seen; the light beam alignment may also change when a new bulb is fitted or if the

mirror is repositioned.

3.2.7.2 Errors due to changes in equipment

The alignment test tool contains no working parts and does not change unless it is damaged. Consequently, the coefficient of variation ascribed to the functioning of the equipment must be zero.

3.2.7.3 Alignment errors due to measurement technique

The measurements are all relative to one another so the only source of error which is not automatically subtracted out is the positioning of the test tool to align with the light beam. The edge of the beam is diffuse and an error of ±2 mm in the zero position is to be expected.

3.2.8 Exposure Time/Tube Current

3.2.8.1 Random Fluctuations

The magnitude of random fluctuations was evaluated using data from a stable system. The standard deviation was 4.96 ms for a mean value of 361 ms giving a coefficient of variation of 1.4%.

3.2.8.2 Equipment

The dosemeter is the only piece of equipment used for this test. It is impossible to separate out the variations due to the timer section of the dosemeter from the variations in the X-ray machine exposure time. It was observed that the first exposure time measurement after the dose meter had been switched on was invariably higher than the remainder which formed a tightly grouped set. To overcome this, the first measurement was always discarded and subsequent measurements used in any calculations. Based on a set of six measurements, therefore with five used for the calculation (n=5) the coefficient of variation of the dosemeter and X-ray unit combined is 0.15%.

3.2.8.3 Measurement technique

The time of exposure measurement is not affected by the measurement technique as long as the ion chamber is subjected to a sufficiently large

exposure to make it trigger. However, when expressed as a percentage, variations are bigger for shorter exposure times due to the limited measurement accuracy of the exposure meter. Therefore it is necessary to use an index which is measured at a fixed mAs (40 mAs at 28 kV is routinely measured). One needs to check that the engineers have not made adjustments to the tube current and also that the system used to estimate this quantity does not drift over time.

3.2.9 Compression

The compression measurements are a safety check and the aim is to ensure that the compression never goes above a critical value. The problem is essentially reduced to one of ensuring that the measurement plus the estimated error is within the limit specified. This way it is highly unlikely that the true value will ever exceed the specified limit of 200N. In practice, the compression was either so far within the limit or so far over it that the accuracy of the measurement never presented a problem.

3.2.10 Radiation Leakage

This is a radiation safety test, the only part of the patient which should be subjected to any radiation is the breast. This test measures the amount of radiation leaking through the X-ray tube housing and checks that it is within the legal limits.

3.2.10.1 Variations in leakage due to random changes in equipment

The leakage measured depend on the output of the tube and the attenuation of the lead shielding within the tube housing. The attenuation is expected to be constant but the output will fluctuate with variations in the kV (CV=1.9%), the tube output (CV=0.4%). A coefficient of variation of 1.9% is expected from these sources.

3.2.10.2 Radiation Leakage variations due to measuring equipment

The equipment used to measure the magnitude of the radiation leakage is the same as that used for the exposure measurements. The analysis in section 3.2.2.2 is therefore used for this section, giving a coefficient of variation of 0.1%

3.2.10.3 Measurement technique

Very low doses are measured at very short distances away from the equipment casing. It is difficult to position the ion chamber reliably and rather difficult to measure the true position with respect to the focal spot because the focal spot lies within the casing. An error of about 1 cm in the measurement of distance of the ion chamber away from the focal spot is expected. Typical measurement distances are 12 cm giving an expected variation of 8.3% due to the inverse square law.

Overall, these combine to give a coefficient of variation of 8.5%. As this measurement is to establish whether the leakage falls below a critical threshold, this large variability is only a problem when the measurement comes close to that threshold. For the equipment tested, the leakage has never approached the threshold.

3.2.11 Field Uniformity

3.2.11.1 Random variations in the field uniformity

Variations of the field strength with position are expected due to a range of geometrical factors. It is possible, but unlikely, that the distribution of the electrons on the anode can vary due to focusing problems, but mainly, variations are due to the statistical nature of the radiation field.

3.2.11.2 Field uniformity variations due to equipment variability

The equipment used for these measurements consisted of 2 mm of Aluminium and a cassette and film. The optical density on the X-ray film is a proxy measure for exposure and could theoretically be altered by film processing, however as long as the densities lie on the straight line portion of the characteristic curve, then the variations in optical density observed are proportional to the exposure, problems arise when the exposures fall either in the shoulder or the toe region of the characteristic curve of the film where the film response is non-linear. If the gradient of the film changes, then the percentage variation in optical density will change in proportion to the gradient.

3.2.11.3 Measurement technique

The measurement technique is very simple and all parts are fixed in position by the geometry of the machine. There can be no variations introduced due to the positioning of the Al filter (so long as it actually covers the beam).

3.2.12 Magnification

3.2.12.1 Random variations

This is determined geometrically and will not change unless the focal spot position has altered (ie a new tube has been installed or the alignment has been adjusted).

3.2.12.2 Equipment

The alignment test tool was used to evaluate the magnification. This has no working parts which can vary, therefore, the coefficient of variation due to this cause must be zero.

3.2.12.3 Measurement technique

The experimental set-up is very much fixed by the geometry of the X-ray unit, the test tool sits on top of the magnification table, the position of the table is fixed by the attachment points to the X-ray unit. The film and cassette are placed in the Bucky. The only variations which can be caused by incorrect experimental technique are:

a) a lateral displacement in the position of the cassette within the Bucky, and
b) a displacement of the position of the test tool in the plane of the magnification table.

Additionally, the measurements themselves have limited accuracy because the edge of the light beam is diffuse and cannot be measured to better than ± 2 mm. The ruler itself has 1mm divisions giving an error of ± 0.5 mm. Combining these together gives an overall error of ± 2.2 mm.

3.2.13 Grid Factors

3.2.13.1 Random fluctuations of grid factors

Changes in grid factor would not be expected unless the grid was damaged or the radiation quality had changed considerably. Consequently, no variability is expected from this source due to random fluctuations.

3.2.13.2 Measuring equipment

The measurement relies on the accuracy of the manual exposure control for making the exposures and also on the performance of the densitometer. The exposure control has been shown to be reproducible, certainly for exposures made on the same day, to better than 0.1%. The densitometer claims to be accurate to within 0.02 units of optical density, i.e. at a density of 1 the combined error is 2%.

3.2.13.3 Measurement technique

The factors most likely to vary are the processor or the film used. As a sensitometric film is usually produced as part of the test procedures, corrections to the film densities can be made if necessary. There may also be inaccuracies introduced due to variations in the intensity of the beam from left to right across the film. The magnitude of these variations are available from field uniformity data where the variation from the centre to the edge is typically 10%, the holes through which the film was exposed for these measurements are positioned much more centrally, the end hole being about 5cm from the centre line. With a field width of 24 cm, i.e. 12 cm from centre to edge, a variation in density due to field non-uniformity of about 4% would be expected.

Overall, the variation expected in grid exposure factor measurements, found by combining the above estimates of error, is 4.5%.

3.3 Optimum Protocol Determined from Results

The aim of the study is that when a set of results of QA tests for an X-ray unit which experienced a tube failure is considered, particular parameters will be identified which indicate the deterioration of the unit. Based on this, and the

expected levels of variability of the measurements calculated in section 3.2, time periods will be identified which enable the changes detected to be considered to be statistically significant.

The mobile mammography unit based at Southend has been selected to represent a unit experiencing tube failure because the data was acquired for nearly two years before the failure occurred. The mobile unit at Barking Havering and Brentwood was selected to represent a situation where the AEC rather than the tube developed a fault, again because the unit developed this fault during the study period but relatively late on. The mobile unit for East London, one of the units which functioned without problems during the period of the study, was used as a control.

3.3.1 The Size of Variations seen Leading up to Tube Failure

Over the period studied there have been three tube failures. The first occurred early on in the data collection programme. The unit was a Siemens Mammomat 2 at Colchester, a static site with relatively low throughput of women screened, the tube failed after approximately 27,000 exposures. As a sequence of only three measurements was made prior to tube failure, it is difficult to draw any conclusions on possible trends.

The second, the Philips Mammodiagnost U-M at Epping, occurred on an old X-ray unit which had previously been used in a hospital before being adapted for screening purposes. Again this occurred early on in the data gathering and makes it difficult to draw conclusions.

The third failure occurred in a dedicated screening unit based at Southend with a throughput of approximately 70 women per day (280 exposures). A sequence of 9 measurements over a period of 18 months are available for analysis. For each parameter which related to the function of the tube, the gradient was calculated and normalised, these results, the fractional change in parameters per day, are shown in table 3.3.1.1 The majority of measurements show little variation over this period, but the output coefficient (output divided by kV^2) and the uncorrected output showed a downward trend. This is shown in figure 3.3.1.1.

An identical screening unit, operating in the Central and East London Service, was used as a control, this unit was also used to provide the

estimate of random errors which might be observed in a system which is in control. There were fluctuations in the output coefficient, but the strong downward trend, which had been seen in the Southend unit, was not observed in the Central and East London unit. In theory, the output varies in proportion to kV^2, so, in order to allow a comparison which does not depend on the tube voltage to be made between the two units, the output coefficient (the output at one metre normalised for kV^2) is used in preference to the uncorrected output. In theory, this eliminates the effect on output of both changes in output due to slightly different kV calibrations between X-ray units, and minor fluctuations in kV from one measurement to the next. The output coefficients from both units have been plotted on the same axes in figure 3.3.1.2

Parameter	Intercept	Gradient	% change per day
Output with kV power relationship	2.20	0.00069	0.031
Output	58.8	-0.033	0.056
kV	27.9	0.00017	0.00061
HVL	0.31	0.000058	0.019
AEC mAs at default	23.8	0.035	0.147
AEC OD at default	1.08	0.0011	0.102
Exposure time	391	-0.11	0.028
Normalised Output	7.52	-0.0042	0.056

Table 3.3.1.1 Change of QA parameters with time on the Southend Basildon and Thurrock mobile

Figure 3.3.1.1

Southend Mobile - Output measurements at 28 kV

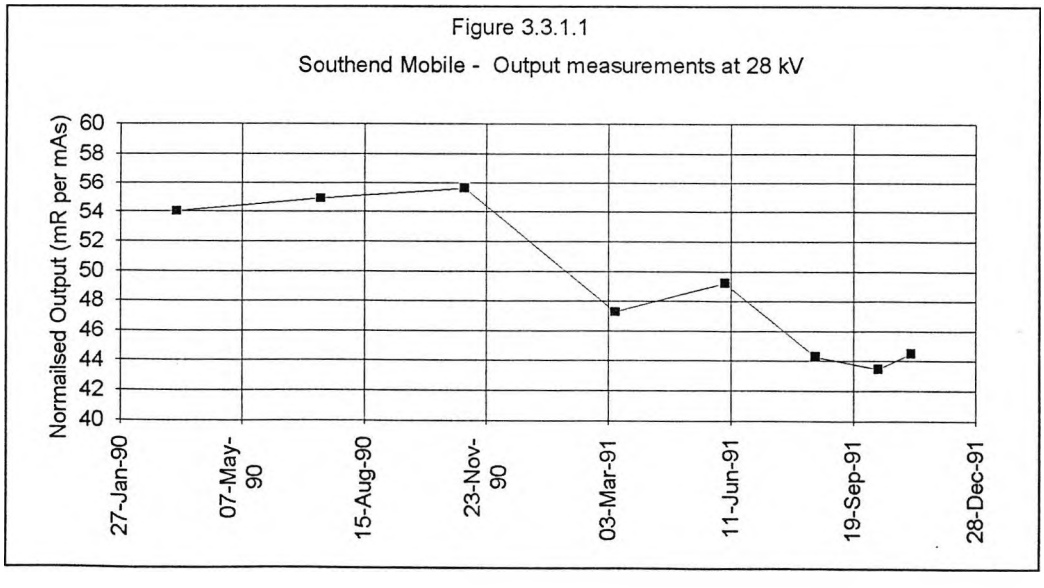
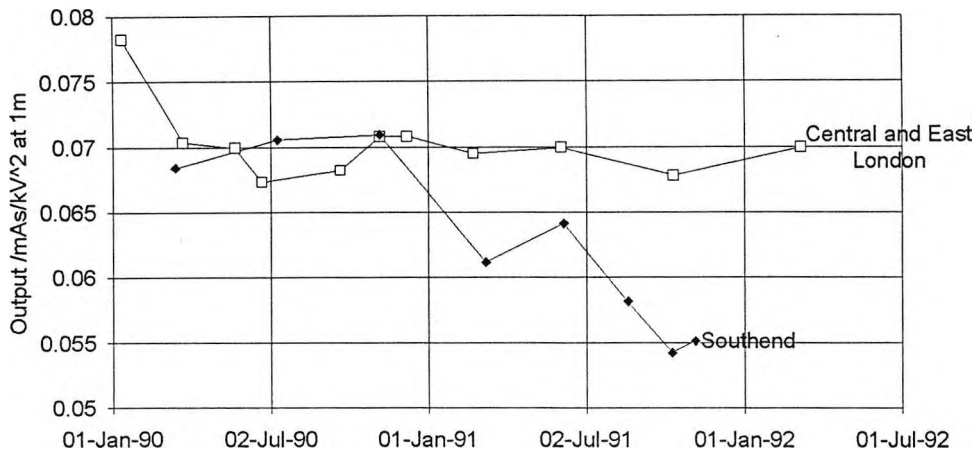


Figure 3.3.1.2

Output Coefficient at 28kV



This corresponds to a change of 0.056% per day in the output of the Southend mobile. It has been assumed that both units start at the same baseline, (which is true within the accuracy of the measurements). As the Central and East London unit has an overall coefficient of variation of 1.9%, it is to be expected that it will take 1.9/0.056 days (=34 days) for output of the gassy tube to reach 1 standard deviation, 68 days to go 2 standard deviations and 102 days to go 3 standard deviations below the baseline. If a normal distribution is assumed, the chance of a reading being greater than 3 standard deviations but still belonging to a normal unit is about 1%. After 102 days (≈3½ months) it should be possible to distinguish between a normal and a gassy tube. The presence of gas within an X-ray tube would otherwise not be inferred until the tube failed. This does assume that the Southend Basildon and Thurrock unit is typical of all gassy X-ray tubes. Without further study it is not possible to say if this is in fact so.

The mechanism by which the gas enters the X-ray tube is not clear, possibilities are:

- evaporation of tungsten from the filament
- evaporation of molybdenum from the target if it becomes overheated
- influx of air molecules via microscopic flaws in the glass envelope

the most likely candidate is the evaporation of molybdenum due to overheating of the anode. The microprocessor limits the size of a single exposure but the build up of heat is not monitored. The thermal cut out due to measuring the oil temperature may be too late to prevent evaporation. With the heavy workloads typical of breast screening, there is a strong temptation to reduce the amount of time allowed for cooling.

3.4 Alternative Strategy

One of the aims of this project is to find the most cost-effective way of performing quality assurance on the X-ray units. The conditions for this are that the image quality is maintained and that any ability to predict tube failure or deterioration should not be compromised. A great many of the parameters which affect image quality also affect the output of the tube so it may be possible to use the output of the tube as an indicator of changes in, for example, filtration or kV. If this works, it means that less measurements need to be made during routine physics testing and raises the possibility of monitoring the tube output via checks on the automatic exposure control, thus enabling a much more "hands-off" approach from physicists. It would be necessary however, to do some periodic checks to cover the aspects of image quality which do not influence the tube output, for example, focal spot size, and to calibrate the post-exposure mAs meter. The possibility of using automatic exposure check depends on the two following hypotheses being true.

Hypothesis 1: a long term falling trend in tube output is indicative of deterioration of the vacuum and tube failure will follow.

Hypothesis 2: if the AEC is stable and the measurement technique is reproducible, then a fall in tube output will be indicated by a rise in mAs when an exposure is made to 4 cm perspex.

Hypothesis 2 was tested by plotting the mAs against time for the Southend Unit (Figure 3.4.1) and calculating the best fit straight line through the points. It has previously been demonstrated that measurements of output coefficient (output divided by kV^2 to remove any tube voltage dependence) showed a fall of 0.04% per day in the period from February 1990 to November 1991 when the tube failed. If the function of the AEC is constant, then the relationship

$$mAs \times Output\ per\ mAs = K \qquad \text{Equation 3.8}$$

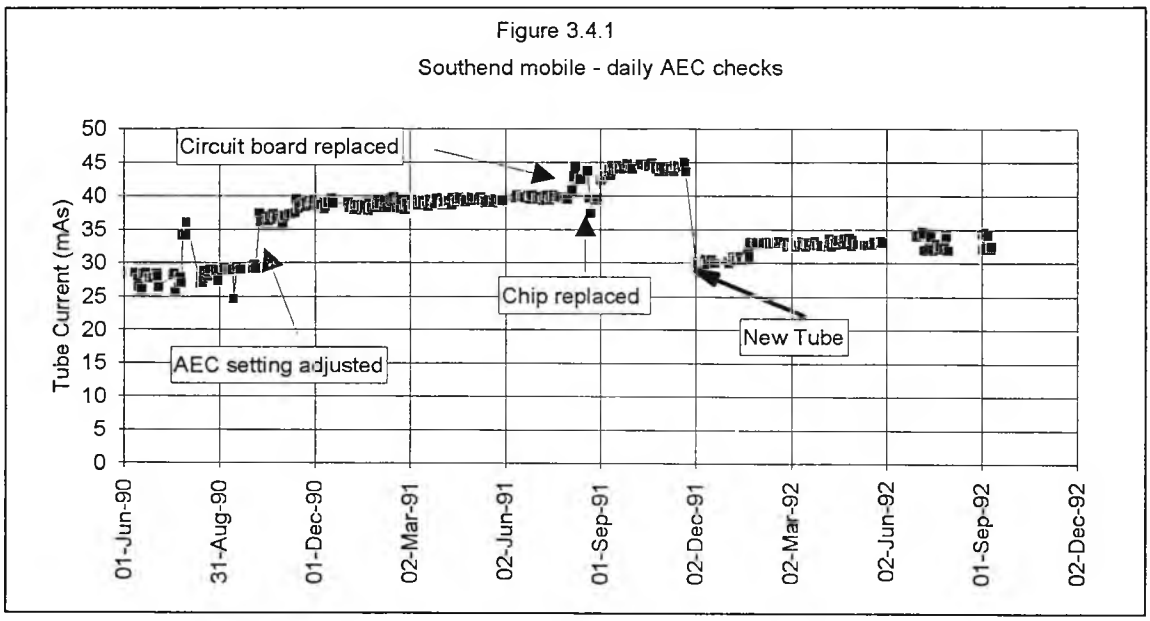
should be true, the constant, K, depends on the optical density selected as the default, the sensitivity of the automatic exposure control and the characteristic response of the film and screen to radiation. It is therefore to be expected that over that same period of time, the mAs recorded will *increase* by 0.04% per day in order to compensate for the falling output.

The step changes seen in figure 3.4.1 are attributable to various interventions

or changes in measurement technique which are indicated on the graph. There is, however, an underlying upward trend, which increases closer to the time of failure. The gradient of the best straight line was evaluated for four separate periods, the final one is the period after the new tube was fitted, the results are summarised in table 3.4.1.

Figure 3.4.1

Southend mobile - daily AEC checks



Calculations show that the gradient of the best fit straight lines are:

From	To	Gradient (mAs/day)	mAs at start of period	Drift (% per day)
12 June 1990	5 October 1990	0.011	27.8	0.04
9 October 1990	1 August 1991	0.008	37.6	0.02
5 August 1991	22 November 1991	0.027	41.9	0.09
2 December 1991	15 January 1992	0.014	30.0	0.07

Table 3.4.1 Drift in mAs measured - Southend mobile

The parameter which is relevant to establishing a correlation between the mAs recorded and the fall in measured output is the rate of change of mAs over the period of study, which can be found from the gradient of the graph. The step changes observed mean that the gradient is relevant only for sections between the step change. In order to evaluate the overall gradient, the gradient for each section was weighted by the time span of the section.

$$\frac{Total\ drift}{No\ of\ days} = \frac{(0.04\% \times 119) + (0.02\% \times 300) + (0.09\% \times 119)}{(119 + 300 + 119)} \qquad \text{Equation 3.9}$$

This gives a drift of 0.043% per day in mAs, very close to the value predicted from the drop in output. This means that the block checks can be used as a proxy measure of the output as long a step changes such as the change of AEC setting are properly accounted for.

3.5 Dose Monitoring

One aspect of quality assurance which cannot adequately be addressed in the above manner is that of dose monitoring. There are legal requirements for an adequate record of dose to be kept, and although physical quality assurance measurements play an important part in such evaluation, the subject requires further consideration.

3.5.1 Requirements

The legal requirements are that adequate records must be kept to enable an evaluation of dose to be made. This is rather vague as it does not specify the accuracy with which the dose needs to be evaluated or the people for whom dose records need to be kept. This could be every woman individually, those suspected to have been overexposed or a value for the population as a whole based on a sample.

The full amount of data which needs to be kept if accurate dose assessment is required for every woman is the kV set (which can be related to the true kV using the physical quality assurance measurements), the thickness of the compressed breast for each exposure and the post exposure mAs for each exposure. With four exposures per woman initially and two exposures per woman on subsequent screens, this quantity of data would be unmanageable and very time consuming to record and store.

Alternatively, thermoluminescent dosemeters (TLDs) could be used to evaluate the entrance and exit doses.

If the accuracy required is reduced, then recording mAs for each exposure gives an estimate of the dose received by the breast tissue. It would be helpful to know what effect ignoring the compressed breast thickness has on the dose estimate.

There is the additional requirement that the dose to the population should be known in order to enable the radiation risk to the screened population to be evaluated and related to the number of induced cancers. This will always be a theoretical estimate since there is no physiological or pathological difference between a naturally occurring cancer and one which has been radiation induced.

3.5.2 Evaluation

In order to address these issues, a survey of doses was undertaken . The kV, post-exposure mAs and compressed breast thickness was recorded for a sample of 400 women for each X-ray unit. Each view was evaluated separately on a spreadsheet in two ways, first, the full data set was used, inverse square law corrections were made based on compressed breast thickness and the ffd of the unit and the conversion factor g (which converts the dose to air into the dose to tissue) was altered according to the compressed breast thickness; second, the compressed breast thickness was taken as 45 mm for all women and the doses recalculated. Conclusions show that :

i The dose for a cranio-caudal view is approximately 25% lower than the dose for a lateral oblique view (averaged over the sample).

ii The dose to the left breast is equal to the dose to the right breast, as would be expected, for two of the three sets of data analysed. The non-matching data cannot be due to the particular make of machine since one of the data sets which does match is from a machine which is identical. It would appear to be an effect either of asymmetrical compression or of using non-matched cassettes.

iii The calculated dose values for the sample are the same whether individual thicknesses are used or and average is assumed. The deviation of the results is greater when an average is assumed.

iv The cranio-caudal doses appear to be approximately 50% greater than the value calculated from physics measurements on 4 cm of perspex.

v The mean compressed breast thicknesses, in mm, for the three units which have been analysed were :

View	Group A	Group B	Group C
Rcc	42	43	48
Lcc	41	42	49
Rlo	46	48	49
Llo	45	48	51

Table 3.5.2 Mean compressed breast thicknesses

which seems to be in reasonable agreement with the 45 mm assumed for IPSM 59.

3.6 Anode Material

When choosing the spectrum of a mammography machine, the argument which was used is that it should be chosen so as to give the best signal to noise ratio per unit dose. [104,105,106,137,138,139] A molybdenum (Mo) target rather than the conventional tungsten (W) which has better mechanical and thermal properties, produces K-lines in the X-ray spectrum at 18 keV and 20.5 keV and heavily absorbs X-rays which have energies below the K edge energy thus producing a spectrum of X-ray energies which fulfil this criterion. The vast majority of X-ray equipment purchased for the UK breast screening programme has been fitted with a molybdenum anode and a molybdenum filter.

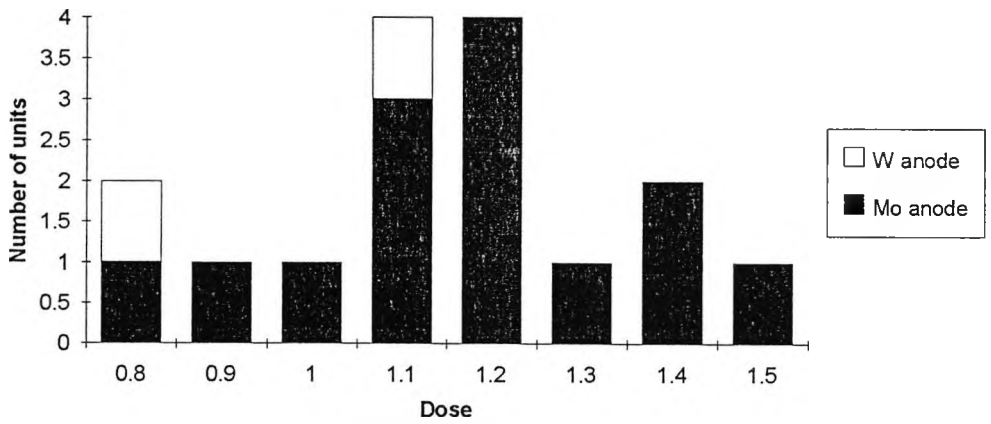
It can also be argued, from the same starting point, that a tungsten anode with a rare earth filter gives a lower dose for women with larger breasts; as these women contribute in a disproportionately heavy manner to the overall population radiation dose, the use of a tungsten anode with suitable filtration should be considered. Aichinger et al [139] reported a comparison between Mo and W anode mammography tubes in which they found that both systems produced acceptable image quality; Jennings et al [106] stated that W anode tube with a K-edge filter was more suitable for mammography than Mo, these results were based on subjective evaluation of the image quality of mastectomy specimens and may well be untypical of the results achieved in practice.

Within NE Thames, W tubes have been used at four centres at various points in time and for varying lengths of time. These will be compared with the other X-ray equipment in NE Thames in terms of mean glandular dose and image quality evaluated using a test object.

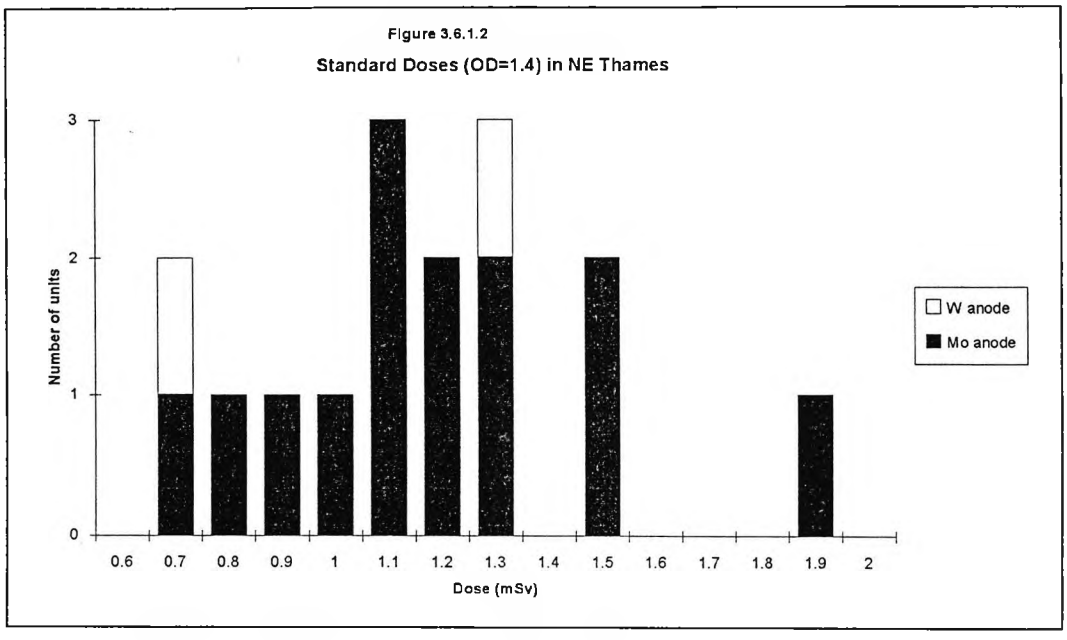
3.6.1 Dose Considerations

The mean glandular dose was evaluated for every piece of equipment in the region during routine quality assurance visits in Spring/Summer 1992, at this point two of the W tubes had been replaced with Mo tubes leaving only two W tubes in use. Figure 3.6.1.1 shows the tubes which had a W target in the anode as white and tubes which had a Mo anode as shaded. It can be seen that the W tubes tend to be at the low dose end of the distribution but that the dose saving is not significant.

Figure 3.6.1.1
Distribution of dose in NE Thames BSP

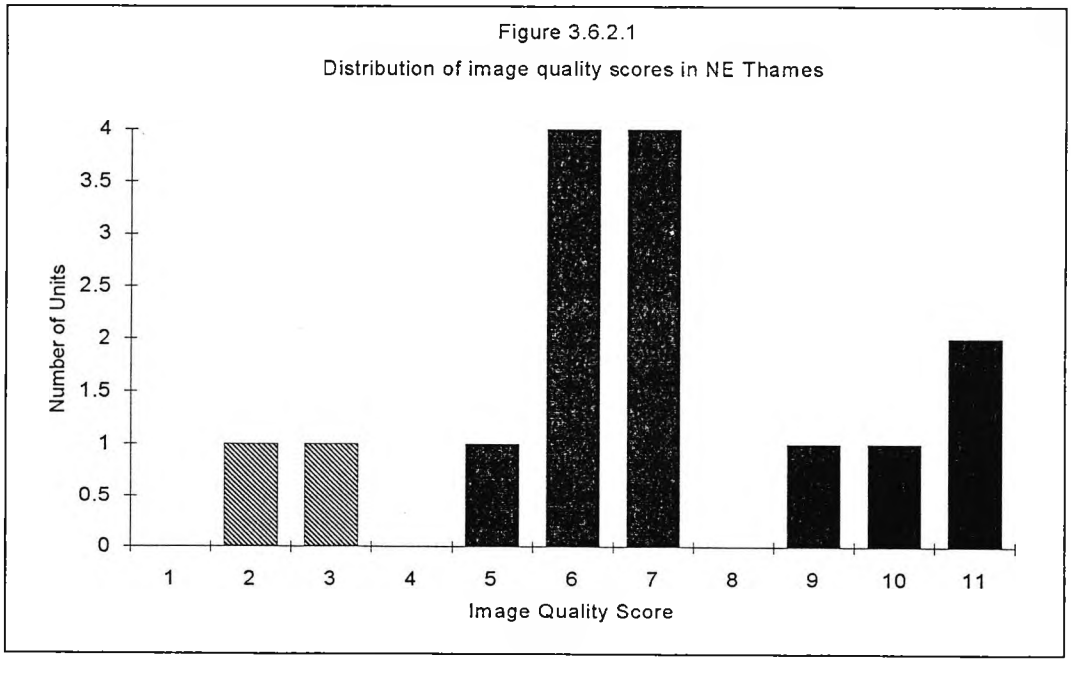


Not all units have the same target optical density (OD) at a voltage of 28kV, some radiologists prefer pale, low density films and therefore have a low target OD, others prefer darker, high density films and have a much higher optical density. In order to make a fair comparison, the actual measured doses have been scaled up or down to give a value representing a target value of 1.4 in OD. These standardised results are shown in figure 3.6.1.2. Again there is no significant difference in the standard dose between tungsten and molybdenum tubes.



3.6.2 Image Quality

The image quality score (the number of objects which were visible compared to the Pritchard standard, as defined in section 3.2.5.2) was recorded for all quality assurance visits. Figure 3.6.2.1 shows the tubes which had a W target in the anode as shaded and tubes which had a Mo anode as solid.



The two tubes with tungsten targets performed less well than those with molybdenum targets, however the clinical significance of this result is uncertain.

3.6.3 Screening Statistics

The true test of image quality is the ability of the system to detect cancers in the screened population. Statistics are available for each of the screening centres, however, X-ray units which had tungsten anodes were invariably used as assessment machines and all of the screening machines had molybdenum anodes, making it impossible to directly evaluate the imaging performance of the different types of tubes. What can be done, however, is to calculate the Bayesian likelihoods (using the model developed in chapter 2) for the assessment stage of the screening process for two units using data from the two corresponding assessment centres, one of which, Southend, used a tungsten anode tube for assessment for the whole of the first round of screening and the other of which, Barking Havering and Brentwood, used a molybdenum anode tube for assessment for the whole of the first round of screening. When studying these results, it is important to remember that the number of women going on to assessment is low, even when added up over a three year period, and consequently the statistical significance of these results will not be high, secondly, that there are many variables which change from one unit to the other not least of which is the processing. Southend used 3M films and chemicals for the assessment stage, whereas Barking Havering and Brentwood used Kodak films and chemistry. Nevertheless, it is interesting to consider how the two set-ups compare (table 3.6.3.1 and 3.6.3.2). The performance during the screening stage appears to be largely similar except for the PPV, although this difference is not statistically significant. At assessment, a small change in specificity can also be observed, but again is not statistically significant.

Parameter	Southend	BHB (Mo)
Sensitivity	99.48%	99.38%
Specificity	96.82%	97.53%
PPV	77.23%	82.56%
NPV	99.94%	99.93%
Accuracy	97.08%	97.73%

Table 3.6.3.1 Statistics for the screening stage of two different units assuming one false negative

The available statistics do not include false negatives as there is a long lead-in time for this information to become available, consequently the data has been calculated twice, once assuming that there is one false

negative outcome from screening for both units and once assuming that there are no false negatives at assessment from either unit. It is reasonable to expect that the number of false negatives at this stage are very low because suspicion has already been aroused and the risk-benefit arguments now favour "playing it safe" and referring the woman to biopsy if there are any doubts at all.

Parameter	Southend (W)	BHB (Mo)
Sensitivity	1.0000	1.0000
Specificity	0.9682	0.9554
PPV	0.7723	0.8256
NPV	1.0000	1.0000
Accuracy	0.9713	0.9780

Table 3.6.3.2 Statistics for the assessment stage of two different units assuming no false negatives

Taking a null hypothesis that the performance of both units is essentially the same, a chi-squared test was performed on the assessment statistics which gave a value of 0.948 (2 d.o.f.) giving $p>>0.1$ which leads to the acceptance of the null hypothesis.

It is unreasonable to assume that only noise affects these results, one would expect to find differences due to the population, the assessment team, the processing and the selected target density for the images, however these, particularly the first two, cannot easily be accounted for. The molybdenum anode tube appears to give a better positive predictive value at the assessment stage, but this result is not of statistical significance.

From this the conclusion has been reached that although molybdenum produces better images than tungsten with the test object, this result does not carry over to the clinical situation where the screening statistics are very similar.

Chapter 4

Control of the Film Processor

4.1 Importance of the processor to image quality

Since the processor is the most variable factor in the imaging chain [140], it is the most critical thing to monitor. All of the processors in North East Thames region have been monitored daily with the measurements being made at the same time each day, as far as possible, in order to achieve maximum reproducibility.

The process control of one unit, that at Whipps Cross hospital, has been monitored for a three year period. The results produced will be used to illustrate various aspects of process control. Full details of the definitions of parameters can be found in appendix C. For the same unit, the daily block checks under automatic exposure control have been saved and a cross correlation between the block densities and the sensitometric parameters was carried out, illustrating the relative influence of AEC and processor on the final density of the film.

In order to understand the results, it is necessary first to consider how the development process works. The formation of the latent image has been explained both by Gurney and Mott [141] and by Mitchell [142], and although they disagree on the exact mechanism, both theories reach the same endpoint; the latent image centre consists of groups of silver atoms in the metallic state within the silver bromide crystals. The speed is partially determined by the crystal size, the larger the crystal size, the faster the film; the gradient of the film is determined in part by the range of grain sizes, a wide range of grain sizes produces a film with a low gradient. The function of the developer is to convert silver bromide crystals which contain a latent image centre to metallic silver and, in an ideal situation, to have no effect on the silver bromide crystals which do not contain a latent image centre. The conversion process is a chemical reduction, the reducing agent used is a superadditive combination of hydroquinone and phenidone, many chemicals can act as a reducing agent on silver bromide, but very few can be selective about which crystals are reduced. Development is very pH dependant and only takes place in an alkaline solution, in order to ensure that the correct pH

is reached, an *accelerator*, either sodium hydroxide or potassium hydroxide, is added to the solution. This would produce a high pH but not in a controlled way; too high a pH will produce overdeveloped films and too low a pH will produce under-developed films. A *buffer* ensures that the solution then stays within a narrow band of that pH, and minimises fluctuations due to pH changes. The activity of the developer is strongly dependant both on the amount of accelerator present and on the effectiveness of the pH buffer.

There is a danger that a highly active developer solution could reduce all of the silver bromide crystals to metallic silver, in which case, the solution would have lost its ability to be selective. It would then produce a final radiograph which was completely blackened and which contained no useful information. In order to prevent this, another chemical called a *restrainer* is added to the developer solution, this may also be called an *anti-foggant*, of which there are two main types, organic restrainers such as benzotiazole and inorganic restrainers such as potassium bromide. The organic type are preferred because they can restrain the level of base + fog (development of unexposed grains) without altering the speed of the system. This is done by increasing the bromine ion barrier around each silver halide crystal, making it energetically difficult for electrons from the developer to react with the silver bromide crystal; where there is a speck of metallic silver in an exposed crystal, this bromide ion barrier is very much reduced enabling development of the crystal to take place. Sometimes, potassium bromide is used as a restrainer, particularly when the speed of the system needs to be reduced, as it does when fresh chemistry is used. Consequently, the starter solution added when chemicals have been replaced contains potassium bromide plus acetic acid whereas the developer replenisher (which is the same as the new chemicals) contains benzotriazole. The other components of the developing solution, solvent (water), sequestrant, hardening agent, wetting agent, anti-frothant and fungicide, do not directly affect the shape of the characteristic curve although they are vital to the correct functioning of the development stage.

4.1.1 Short term variations during batch processing

A characteristic of the breast screening programme that does not generally occur in diagnostic radiography is the need to process a very large number of films in a short space of time. This happens because,

on a mobile unit, where there is no processing available on board, which is normally the case, the films exposed on a mobile unit need to be processed at the end of the working day. Consequently the processor at the hospital base is free to develop the films which are being taken at the static site as part of the second stage assessment during the normal working day.

A study was done to see how a processor coped with batch processing. One hundred films were exposed under AEC, a sensitometry strip was printed on the rear edge which had been shielded by a lead strip during the X-ray exposure and the films were processed.

Two closely matched cassettes were used alternately in order to allow the exposures to be made quickly enough to keep the processor working at its maximum speed. This means that the experiment was simulating the way in which the processor works when a batch of films is being processed, with one 24x30 cm film being processed approximately every 70 seconds. Additionally this meant that the X-ray unit was working at a similar rate to that encountered during a busy screening session where four films will be taken per woman and appointments can be at four to five minute intervals, the main difference being that in a screening situation, the exposures are not evenly spaced.

The optical density of the film exposed under 4cm of perspex showed a downward drift in optical density over the course of 100 films (figure 4.1.1.1) amounting to a total change of 0.07 optical density units from a starting density of 1.21. This change would not be detectable by the eye but indicates a drift either in the processor or the automatic exposure control or both.

The parameters which describe the characteristic curve of the films, Dmin, speed, gradient and Dmax show negligible variation during the course of the experiment except for the speed (table 4.1.1 and figures 4.1.1.4 and 4.1.1.5). A representative of the film manufacturer commented that this was "quite a large change" and suggested two possible causes, the replenishment rate may have been too low, or,

addition of replenisher caused the temperature in the developer tank to fall, the latter would not have caused a problem had the films been processed at a lower rate because the temperature stabilisation would have had sufficient time to reach equilibrium.

	Dmin	Speed	Gradient	Dmax
Gradient	-1.5 E-5	-0.156	-3.8 E-5	-4.6 E-4
Mean value	0.188	303	4.91	3.34
Standard Deviation	0.004	12	0.03	0.02
Coefficient of variation	2.13%	3.96%	0.61%	0.60%

Table 4.1.1: Drift in sensitometry parameters during batch processing

The automatic exposure control performance could also have contributed to the observed drift in optical density. Its performance over the course of the test is shown in figure 4.1.1.2, if the output of the tube remains constant and the AEC is functioning in an ideal manner, the graph would be a horizontal line. Figure 4.1.1.2 shows the AEC response to be divided into two parts, the initial stage during which the recorded mAs values do indeed form a horizontal line with the usual statistical variations which lasts for approximately 60 films, and the remaining 40 films for which a small but definite downward trend can be observed. If these changes are due to the radiation detection circuitry alone (in other words if the AED was the sole cause of the change), a scatter plot of optical density against mAs would give a straight line whose gradient represents the change in optical density per unit change in mAs and is consequently related to the performance of the intensifying screen in the cassette and the characteristic curve of the film. Figure 4.1.1.3 shows a correlation of 0.577 ($p<0.001$ for $n=100$) between optical density and mAs and seems to be composed of two straight lines of different gradients which partially overlap, suggesting that the observed drift was due to the AED. From this, one may infer that the output per mAs must have been constant.

Figure 4.1.1.1

BHB batch processing trial - Variation in Optical Density through the trial

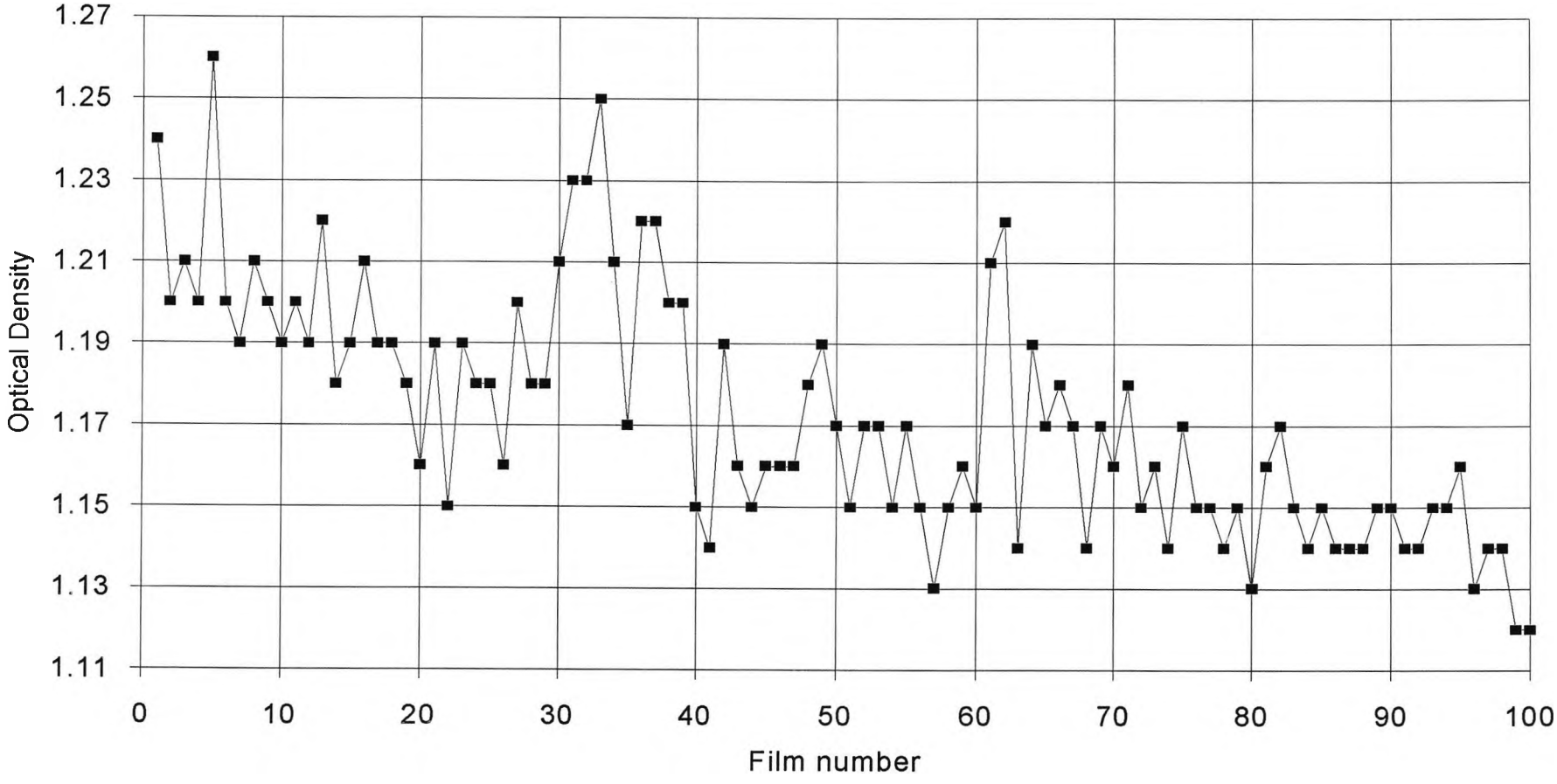


Figure 4.1.1.2

Barking Havering and Brentwood AEC consistency trial

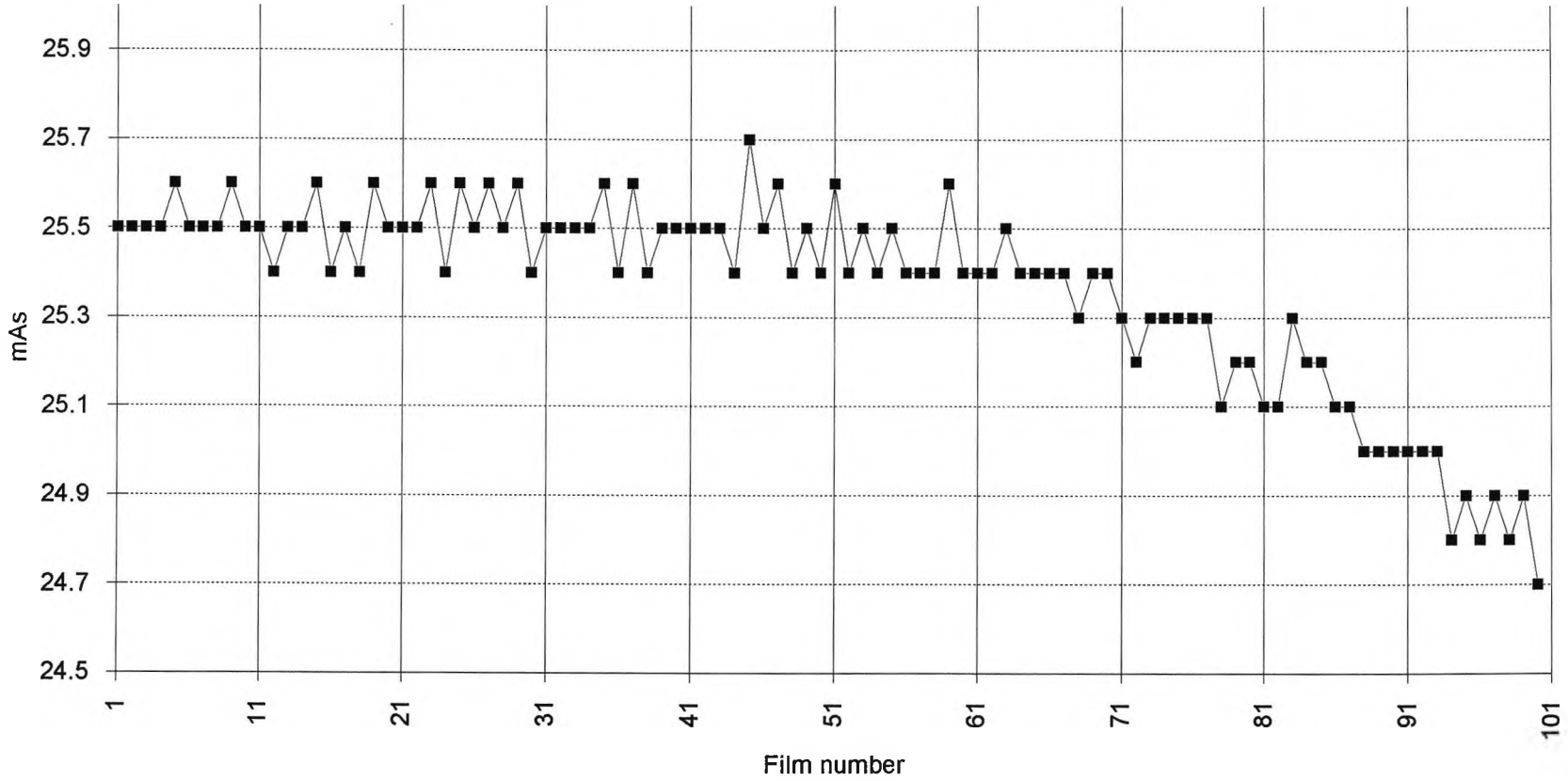


Figure 4.1.1.3

Consistency trial

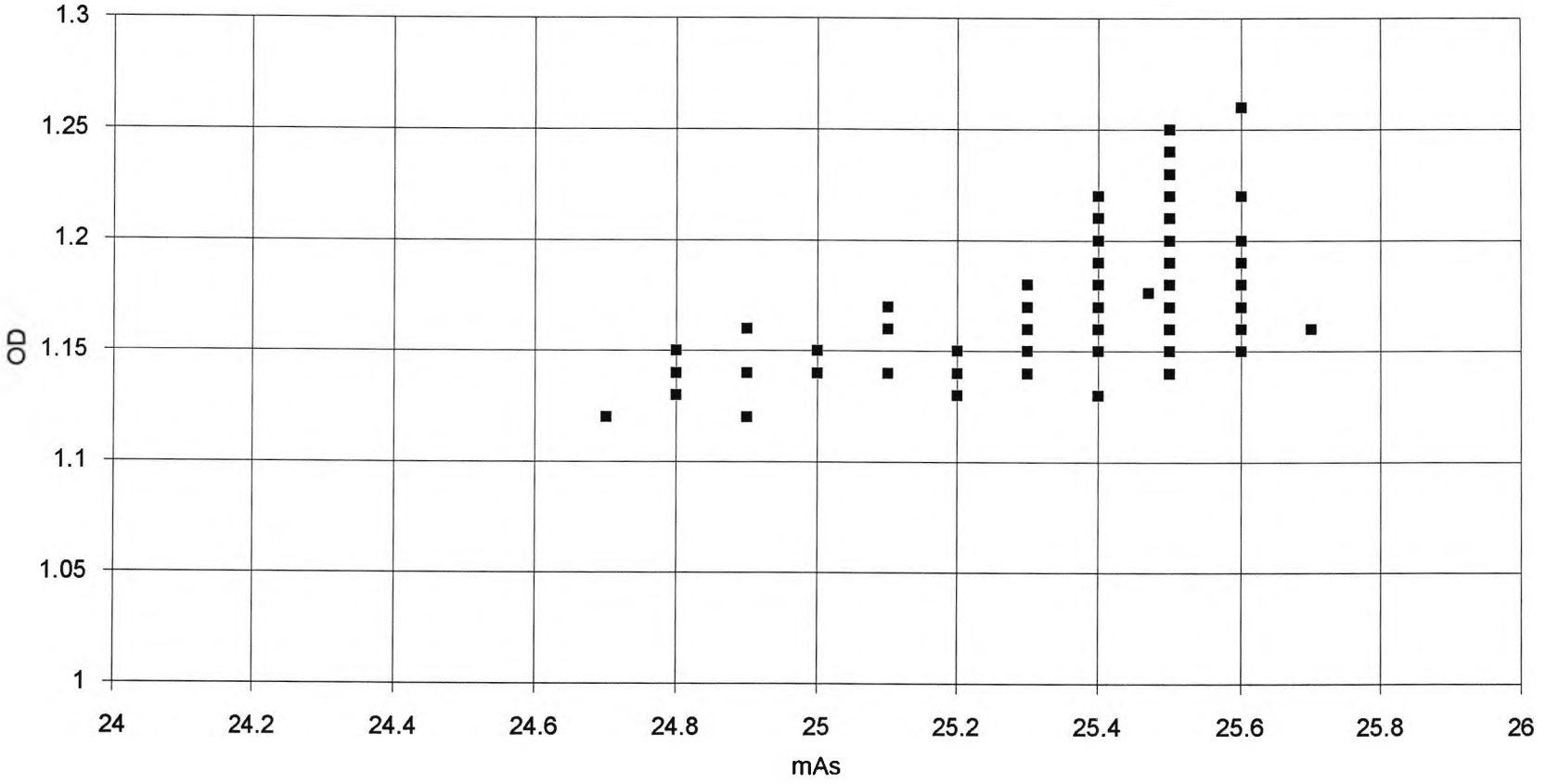


Figure 4.1.1.4

BHB batch processing trial

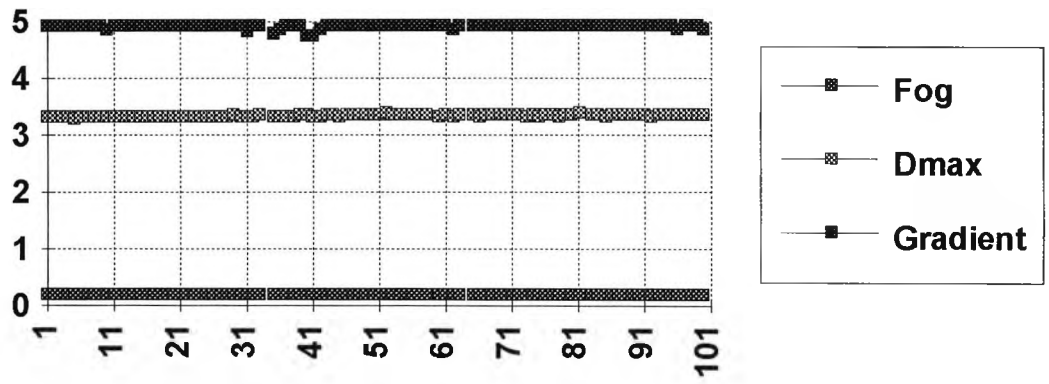
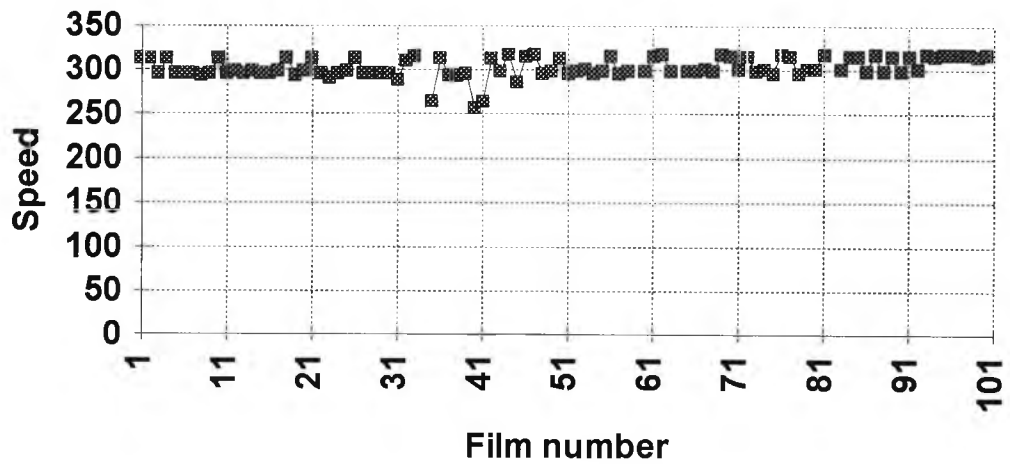


Figure 4.1.1.5

BHB batch processing trial



4.1.2 Medium term variations, between services

In general, a processor will stay in statistical control if the workload does not change substantially and no untoward events happen such as the thermostat developing a fault, chemicals becoming contaminated or the water supply failing. One event which does occur regularly over the life of a processor is the replacement of chemicals at a routine service. This section looks at how this routine procedure affects the function of the processor and what precautions, if any, are needed to maintain the image quality achieved when the processor is in control.

The effect is the opposite of what would be expected according to conventional processing expectations. Generally, fresh chemicals are most active, and as films are developed and the active ions are consumed, the chemistry becomes depleted and the activity falls. In order to maintain a reasonable level of activity, some fresh developer replenisher is added to the developer tank and fixer replenisher to the fixer tanks; because of mixing within the tank, the original level of activity is never achieved and eventually a new equilibrium is reached at which the processor will remain unless some event causes it to deviate.

In mammography, the film has only one layer of emulsion in order to increase the sharpness of the image. The film would tend to be slow (have a low sensitivity to X-rays) compared to normal double-sided X-ray film were it not for the addition of chemical accelerators which are incorporated in the emulsion. These speed up the development process and allow good images to be produced at a low radiation dose. As films are processed, a small amount of this accelerator is leached into the developer causing the chemistry to become more active. Two opposing processes are occurring, the developer is being depleted but at the same time, the accelerator is making the remaining developer more efficient. Addition of replenisher replaces a proportion of this mixture which maintains the concentration of the developer and restrains the accelerator at the same time. The net result is an equilibrium which produces films somewhat faster than when completely fresh chemistry is used. [143]

Figures 4.1.2.1 to 4.1.2.4 show the parameters fog (Dmin), film speed, film gradient and maximum optical density (Dmax) for a six month period from 2nd April 1991 to 29th September 1991. The dates on which fresh chemicals were added are marked on these graphs. It would be expected that the film speed will show the most marked variation with changes of chemistry, although the fog might also be expected to react to changes in the developer.

Figure 4.1.2.1

Whipps Cross - April '91 to September '91

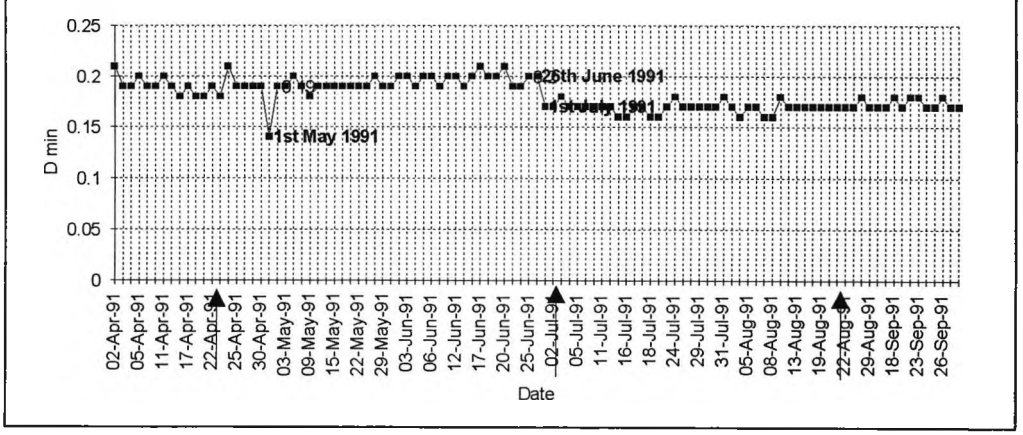
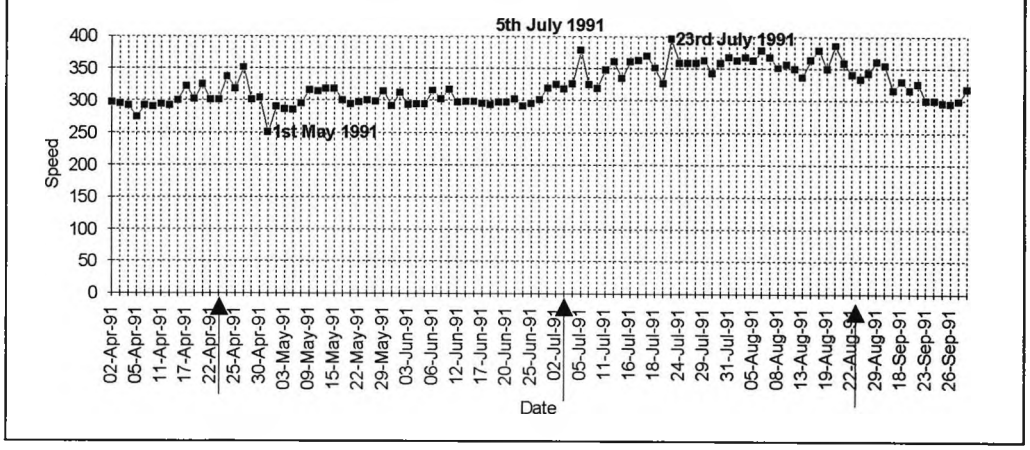
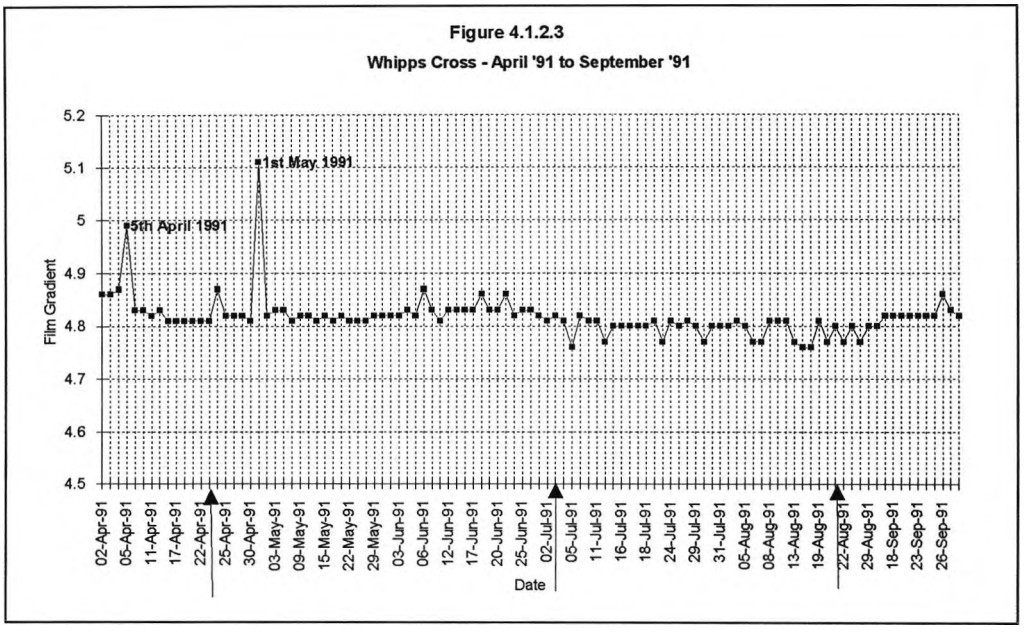
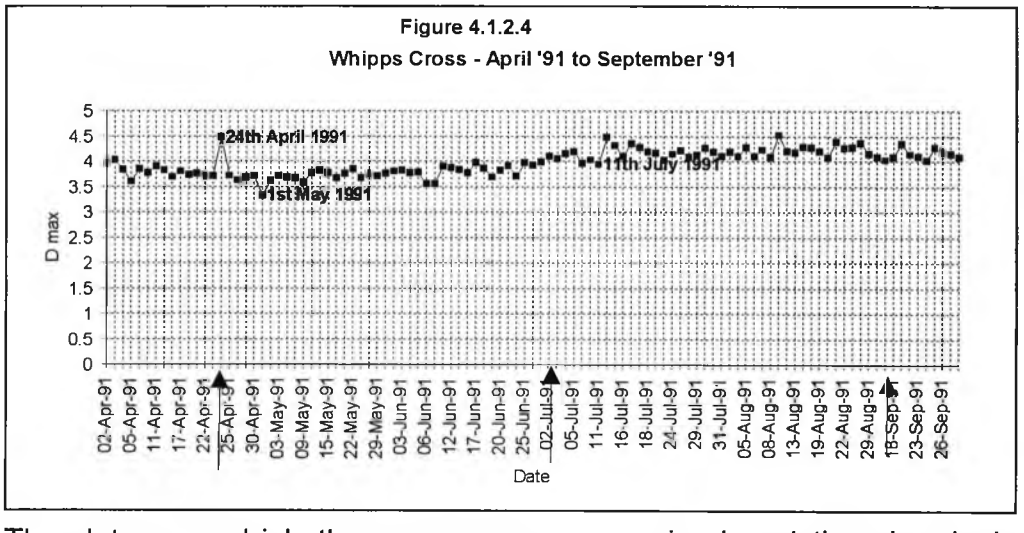


Figure 4.1.2.2

Whipps Cross - April '91 to September '91







The dates on which the processor was serviced and the chemicals replaced have been marked on each of the processor control charts with an arrow. In addition to fresh chemicals being added at the time of a service, there have been other occasions on which the developer and/or replenisher have been replaced, for example when the developer becomes contaminated with splashes of fixer, these are also marked on the control charts with an arrow and listed below in table 4.1.2. From the graphs it seems that on the 1st May 1991, something happened which caused a noticeable excursion in all of the measured parameters, however, nothing in the QA record corresponds to this; the excursions appear to be random in nature except for the increase in Dmax on 24th April which corresponds with the replacement of the processor chemicals following the failure of the water supply.

Services	New Chemicals
13th March 1990	20th February 1990
16th July 1990	23rd March 1990
27th November1990	12th September 1990
22 March 1991	23rd October 1990
28 June 1991	17th January 1991
1st November 1991	25th January 1991
6th March 1992	24th April 1991
3rd July 1992	2nd September 1991
6th November 1992	19th November 1991
12th February 1993	17th December 1991
16th July 1993	20th October 1992
10th October 1993	

Table 4.1.2: Dates of replacement of chemicals in the processor.

This raises two questions:

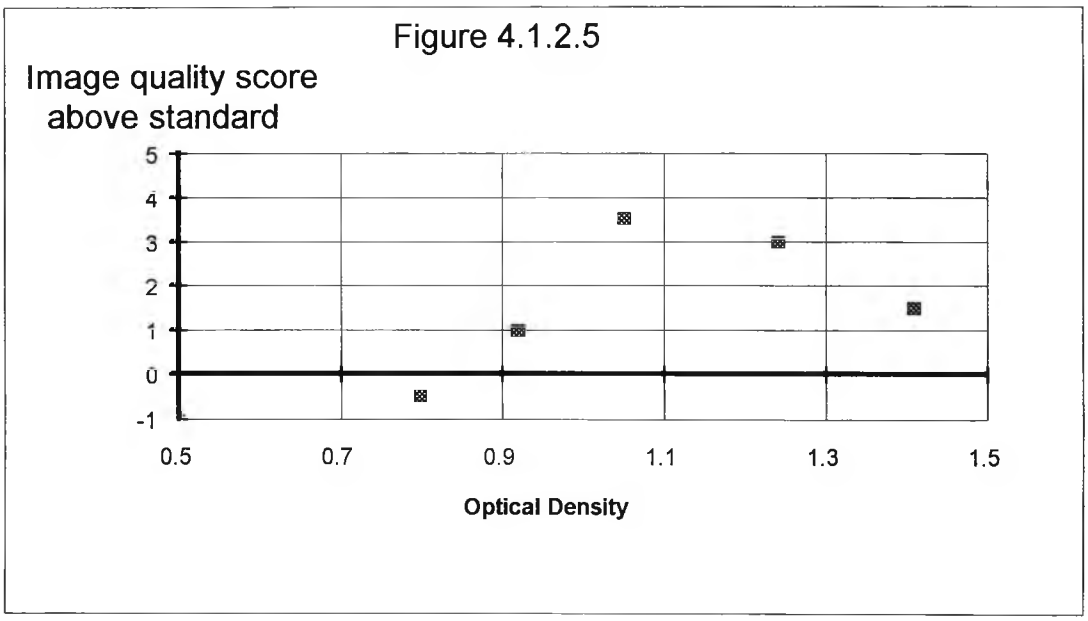
i Will this change of film parameters cause any loss of image quality ?

ii Can anything be done to counteract this short term effect ?

In order to address the problem in i), the following test was performed. A series of image quality films were produced using the same kV throughout and a range of exposure factors so that the mean density of the film would change whilst everything else remained constant. this is equivalent to changing the film speed and keeping the fog, film gradient and Dmax constant, the larger exposures produced darker films which were equivalent to the standard exposure with a faster (more sensitive) film. The resulting images were scored by a physicist and the score plotted as a function of optical density measured at a fixed point on the film stepwedge. The graph in figure 4.1.2.5 shows that there is an optimum density at which the image quality score reaches a maximum value. this agrees well with a similar experiment

by Murray et al [64] where the optical density was adjusted by making copy films of the original test film.

Figure 4.1.2.5



If it is assumed that the imaging process has been optimised, i.e. that the mean density of a clinical film corresponds to the images produced when the optical density at the test point of the image quality test films was 1.0, then any change in film speed which does not simultaneously alter the slope of the characteristic curve will cause a drop in the image quality.

The peak of the image quality/optical density plot is broad, a change in image quality score of 1 can be treated as negligible, being well within the range expected for intra-observer variation; consequently a density range of 0.95 to 1.3 is acceptable insofar as the changes which it causes are not sufficient to be noticeable to the observer. To translate this into speed terms, it is necessary to make the assumption that the measurements are made on the straight line portion of the curve, which they are (although the mechanism for loss of image quality probably involves the parts of the image which appear in either the shoulder or the toe of the curve, this explains the asymmetry in the fall off of image quality).

First, let the optical density of 1.0 recorded at the point of highest image quality score correspond to a film speed of 300; this is purely

arbitrary, the actual density produced depends not only on the film speed but on the X-ray exposure factors too. The minimum optical density which will produce no loss in image quality is 0.95, so if the relationship between film gradient, density and exposure is assumed to be

$$\gamma = \frac{OD_2 - OD_1}{\log_{10}(E_2) - \log_{10}(E_1)} \qquad \text{Equation 4.1}$$

where γ is the gradient, OD refers to the optical density at the top and bottom of the straight line portion of the characteristic curve and E is the exposure needed to produce those densities.

then the change in speed can be found by putting

$$\log_{10}(E_2) - \log_{10}(E_1) = \frac{OD_2 - OD_1}{\gamma} \qquad \text{Equation 4.2}$$

so taking γ to be 4.8, which is typical of the Kodak MinR-E film, and taking logs the equation becomes:

$$\frac{E_2}{E_1} = 10^{0.05/4.8} \qquad \text{Equation 4.3}$$

or a factor of 1.02 decrease in exposure. This is equivalent to a drop in speed to 300/1.02 = 293

Similarly the limit set by the upper limit of optical density of 1.3 gives a maximum speed, which would not *noticeably* affect the image quality, of 346. Assuming that the system was indeed optimised, the asymmetry of the image quality to optical density relationship means that small decreases in speed are unacceptable whereas relatively large *increases* in speed, up to 6.6 times the acceptable drop in speed, still produce images with no noticeable degradation.

Looking at the control data, the film does go outside of these limits from time to time. The limits determined in this way are the specifications required to ensure consistently good image quality, the control limits are statistically derived and give a standard deviation in

speed of 17 speed units for the Kodak film initially, 28 speed units for the Kodak film second stage and 15 speed units for the DuPont film used after 19th June 1992. It is apparent that using the standard statistical results that 95% of the data will lie within 2 standard deviations of the mean, and that 67% will lie within 1 standard deviation of the mean, that if the processor is statistically in control, over five percent of the time the results will be outside of the desired specification, even for the best case. In order to achieve the objective of no noticeable deterioration in image quality due to the film, screen and processor, the system would need to be modified.

In answer to the second question, it is necessary to distinguish between random variations which will occur anyway, and about which nothing can be done without redesigning the system, and variations produced by external actions such as the replacement of chemicals in either the developer tank or the processor tank. One strategy for dealing with new chemicals is to reserve some of the old chemicals prior to a service and mix them in with the new chemicals when the tanks are refilled. Such a procedure would be unsuitable if the chemicals were being changed because of contamination or some other fault which has caused the system to go out of control. In that type of situation, an alternative strategy is needed.

It is normal practice for standard processing to add a "starter" solution to fresh chemicals because the replacement chemicals used are identical to the replenishment solutions. The starter serves to reduce the pH of the developer, thus making the chemicals less active, and to provide additional Bromine ions which would normally be released from the film during processing. This has the effect of bringing the solution close to the final equilibrium right from the start. In mammography, this is counter productive, the film emulsion itself leaches an accelerator during development, so as time goes on, an equilibrium is reached where the speed of the process is faster than with fresh chemicals, conventional starter solutions would make the problem worse [144].

The effect was first noticed in mammography at the Jarvis clinic in 1989 but has also been observed where other single sided emulsions

are used, 3M ultrasound film in 1987, Fuji mammographic film in 1989, Kodak nuclear medicine film in 1991 and Fuji industrial film in 1991 [145].

After a change of chemistry therefore, the speed of the system falls and the contrast rises in the first instance, exactly the opposite effect of that seen in duplitised (double sided) films. As replenisher is added either the accelerator is leached from the emulsion or the release of iodide during the development process occurs, or both. In either case, the speed begins to rise and the contrast to fall until an equilibrium is reached. The point of equilibrium and the time taken to reach it depends on the replenishment rate and the throughput of film. As a rule of thumb, it can be expected that equilibrium will be reached when the amount of replenisher added is equal to three times the volume of the developing tank. As a consequence, a specific mammographic starter has been developed which accelerates rather than restrains the development process and starts the solution off at a point much closer to the final equilibrium.

Looking at figures 4.1.2.1 to 4.1.2.4 (pages 150-151) it is evident that this strategy is highly successful, the changes in sensitometry with fresh chemistry are smaller than random changes.

4.1.3 Long term variations, three years of sensitometry

The processor may vary over an even longer time scale, this long term variation might be due to factors such as wear of the valves in either the input of chemicals to the processor or the output of chemicals from the automixer which would allow the replenishment rates to drift over a long period of time, or variations in components which control for example, the temperature of the developer. Variations of any significance are measurable, and will therefore show up on long term charts of processor parameters. Figures 4.1.3.1 to 4.1.3.4 show the long term variations in D_{min} , speed, slope and D_{max} of the film at one particular centre.

Figure 4.1.3.1 Variation in Base + Fog
Whipps Cross

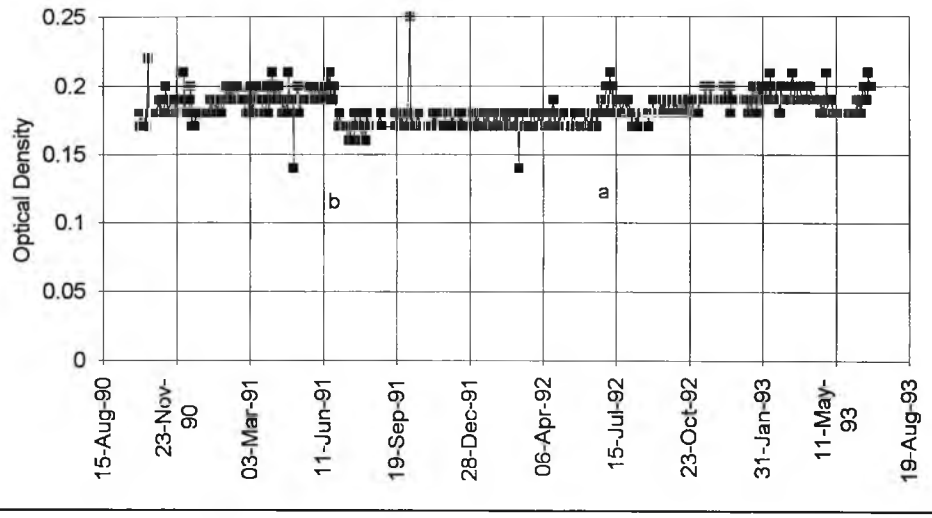


Figure 4.2.3.1 Variations in film speed
Whipps Cross

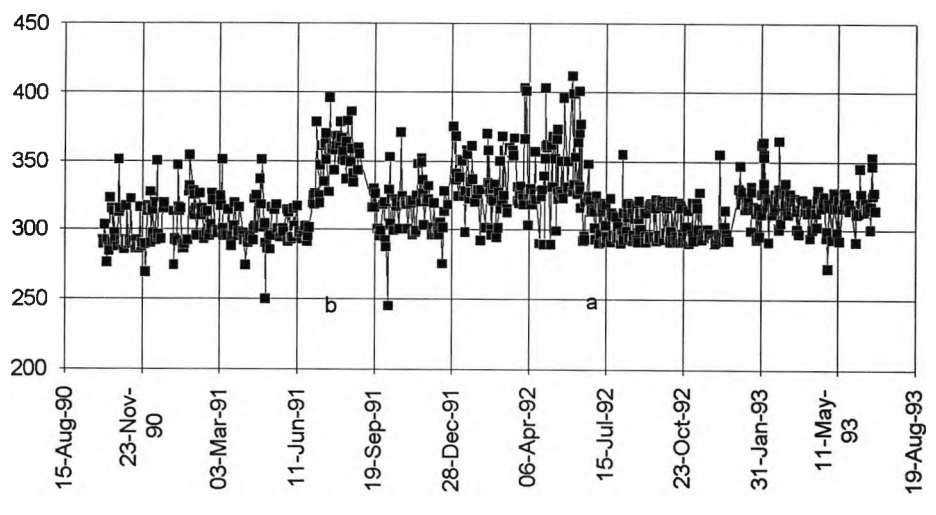


Figure 4.1.3.3 Variations in film gradient

Whipps Cross

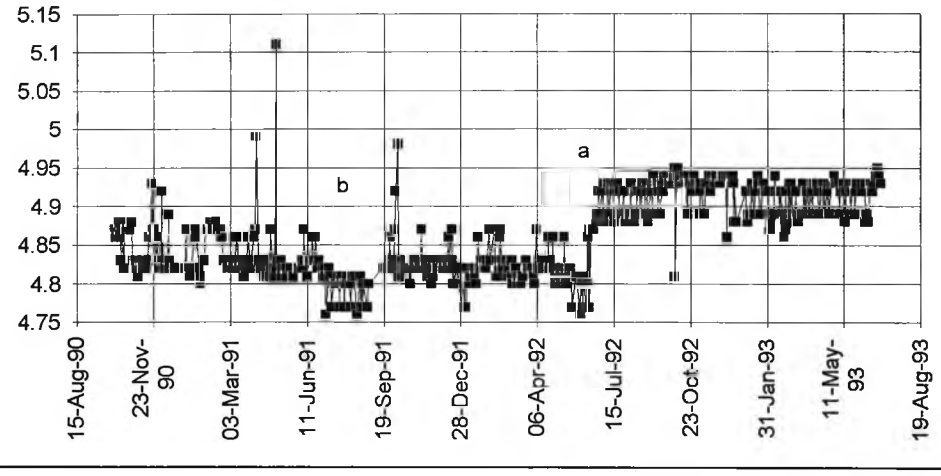
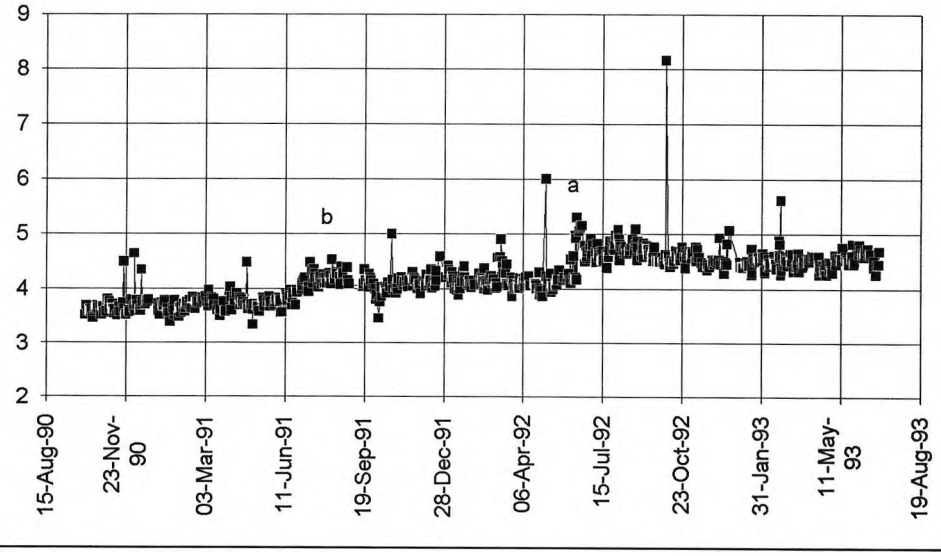


Figure 4.1.3.4 Variation in Dmax

Whipps Cross



The change from Kodak film on 18th June 1992 to DuPont film on 21st June 1992 is clearly seen (a). A change taking place on 17th June 1991 can also be seen (b). A Kodak representative has been approached to see if the film had changed in any way but according to their records there should be no difference from earlier results, another possible explanation is that the replenishment rates were changed on that day, however, according to the records, no such change in replenishment rates occurred. This change can be seen in the slope and the Dmax of the characteristic curves as a step change. The excursions away from the usual values which occurred on 1st May and

7th October 1991 affected all four senistometry parameters;. Excursions involving two parameters were noted on 23rd October 1991, 2nd October 1992 and 24th February 1993; there were no incidents recorded on these dates which would explain the changes. It appears that the majority of the excursions seen have no special cause and are simply random. The Dmax value is particluarly prone to wide fluctuations because the Dmax was not actually measured it was extrapolated from data producd by a 9 step sensitometer; the downward excursions appear to have some significance (they correspond to excursions in other parameters) whereas the upward excursions appear to be artefacts of the extrapolation process. The other point to note is that isolated values are uninformative; a value which is an extreme excursion at one point in time may be within the normal range a year later. Any interpretation needs to be done with regard to the general values and trends in a period of two weeks up to the day in question.

For the purposes of this analysis, the statistics have been done separately on the three groups of film as described in section 4.1.2.

	Kodak 1	Kodak 2	DuPont
Mean	303	333	312
Standard deviation	17	28	15
Coefficient of variation	5.6	8.3	4.8
Gradient of best straight line fit (drift)	0.02	0.02	0.04

Table 4.1.3.1 - Process Control Statistics for film Speed

It is informative to consider whether or not these kind of coefficients of variation which occur over a period of about twelve months are of the same kind of order as those produced when films are batch processed. The coefficient of variation of each parameter is shown for a batch of 100 films and also for the three groups of data from Whipps Cross in table 4.3.2.

From table 4.1.3.2, it can be seen that the variations of processor parameters seen during routine processing are generally much greater than their counterparts in batch processing, with the exception of the

film gradient, which shows a similar level of variation for all four groups of film. The data from the batch processing experiment all comes from a single batch of film and on that basis alone would be expected to be more consistent than the other measurements. It should also be noted that due to initial problems with speed variations in the DuPont film, the batches used were pre-selected for consistency of speed so the results should also be significantly less variable.

	Batch Processing (Kodak)	Kodak 1	Kodak 2	DuPont
Dmin	2.1%	5.1%	5.1%	3.9%
Speed	3.9%	5.6%	8.3%	4.8%
Gradient	0.6%	0.7%	0.6%	0.5%
Dmax	0.6%	4.7%	6.4%	5.8%

Table 4.1.3.2 - Coefficient of Variation of Processor Parameters - Drift over time

4.2 The influence of the processor on X-ray images.

The assertion has been made that the processor strongly influences the final outcome in imaging. This has been examined in two ways, the optical density of the block checks has been used as an indicator of mean density achieved on the film, and the image quality score has also been used as a measure of the influence of the processor.

4.2.1 Mean film density

The optical density of the block checks has been plotted against the various processor parameters to see how each parameter influences the measured optical density. The OD has been plotted against speed both for a batch process (figure 4.2.1.1) and also for the daily block checks (figure 4.2.1.2) the OD was also plotted against gradient for the daily block checks (figure 4.2.1.3) and OD against fog (figure 4.2.1.4). These show very little correlation between speed or gradient of the film and the optical density of the block check films.

4.2.2 Image quality

The image quality score for routine physics QA visits has been compared with the various processor parameters. The image quality number denotes the number of objects seen compared to the minimum acceptable standard defined by the Pritchard report [3]. So if there are eight 6mm circles visible, and the Pritchard standard is seven, a score of +1 is achieved for the 6mm circles, the same procedure is carried out for the other items specified in Pritchard and the scores for all of the items added together. A positive score therefore denotes a system which is better than the Pritchard standard and a negative score denotes a system which is below the Pritchard standard, this scoring method does however have the drawback that a good score on one aspect of the test object may obscure a feature for which the imaging system is not up to standard. Table 4.2.2 shows the image quality scores and the corresponding sensitometry values.

Figure 4.2.1.1

BHB batch processing trial

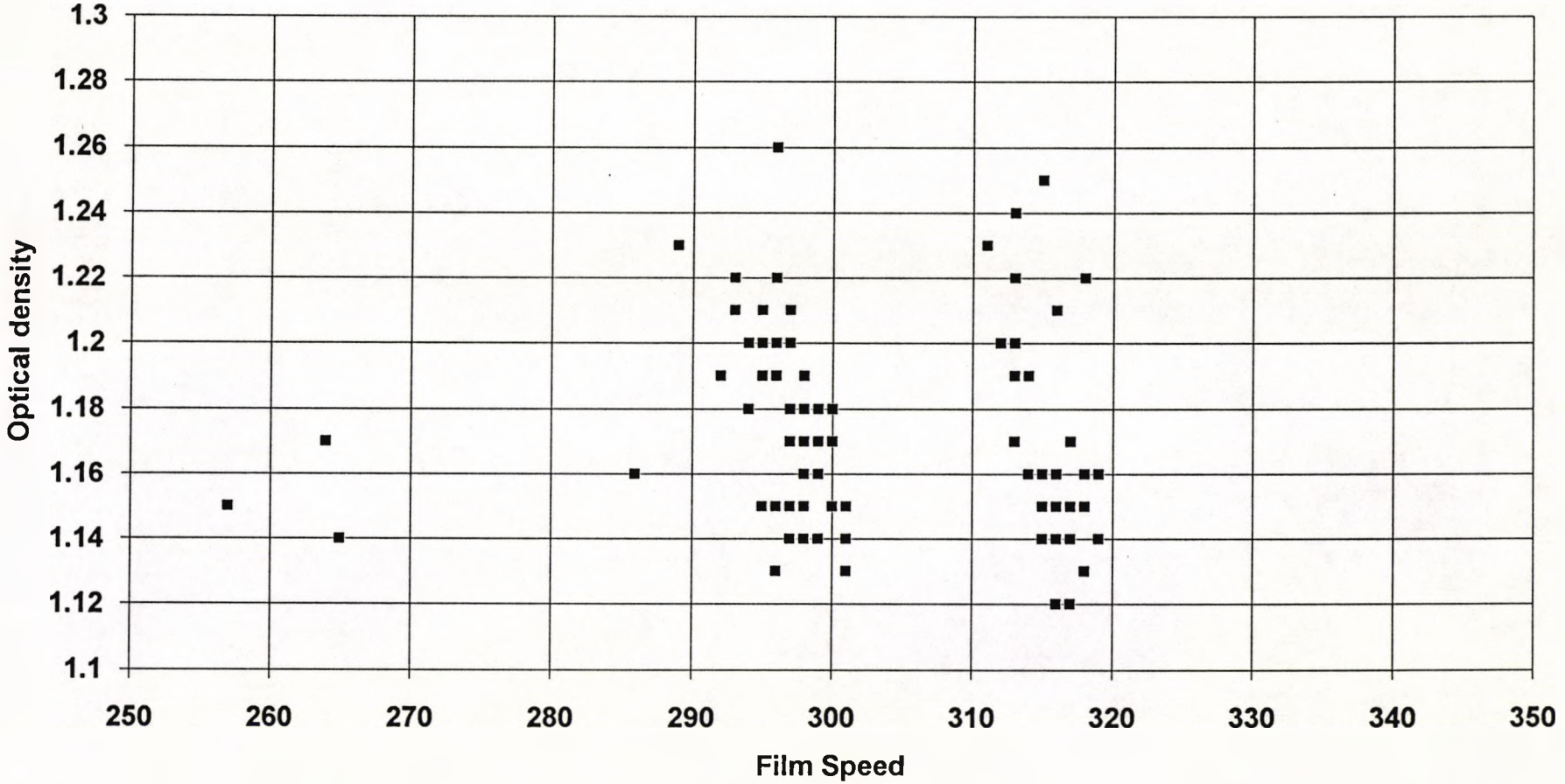


Figure 4.2.1.2

Relationship between film speed and daily block check denisty at Whipps Cross

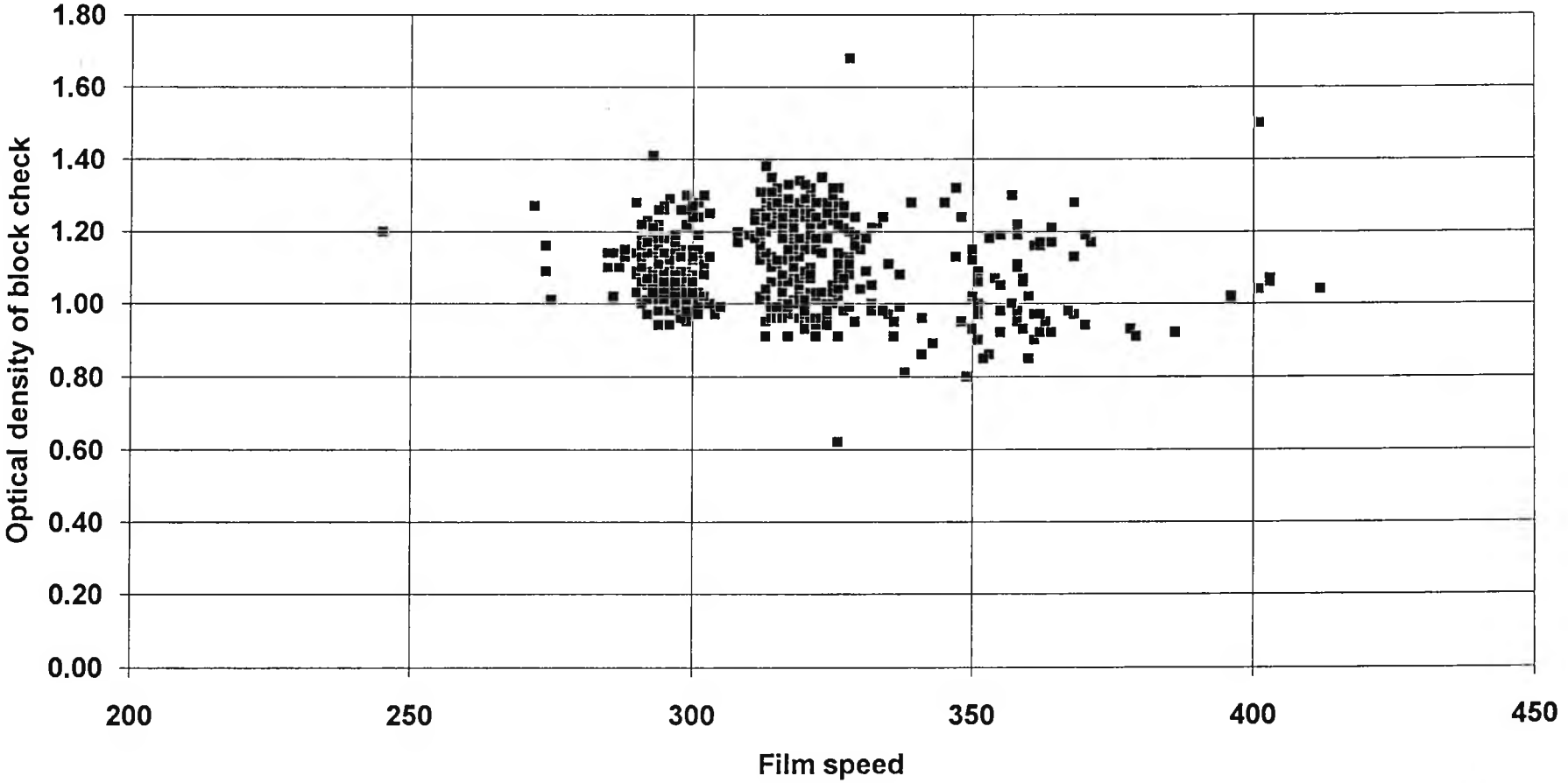


Figure 4.2.1.3

Relationship between film gradient and daily block check density at Whipps Cross

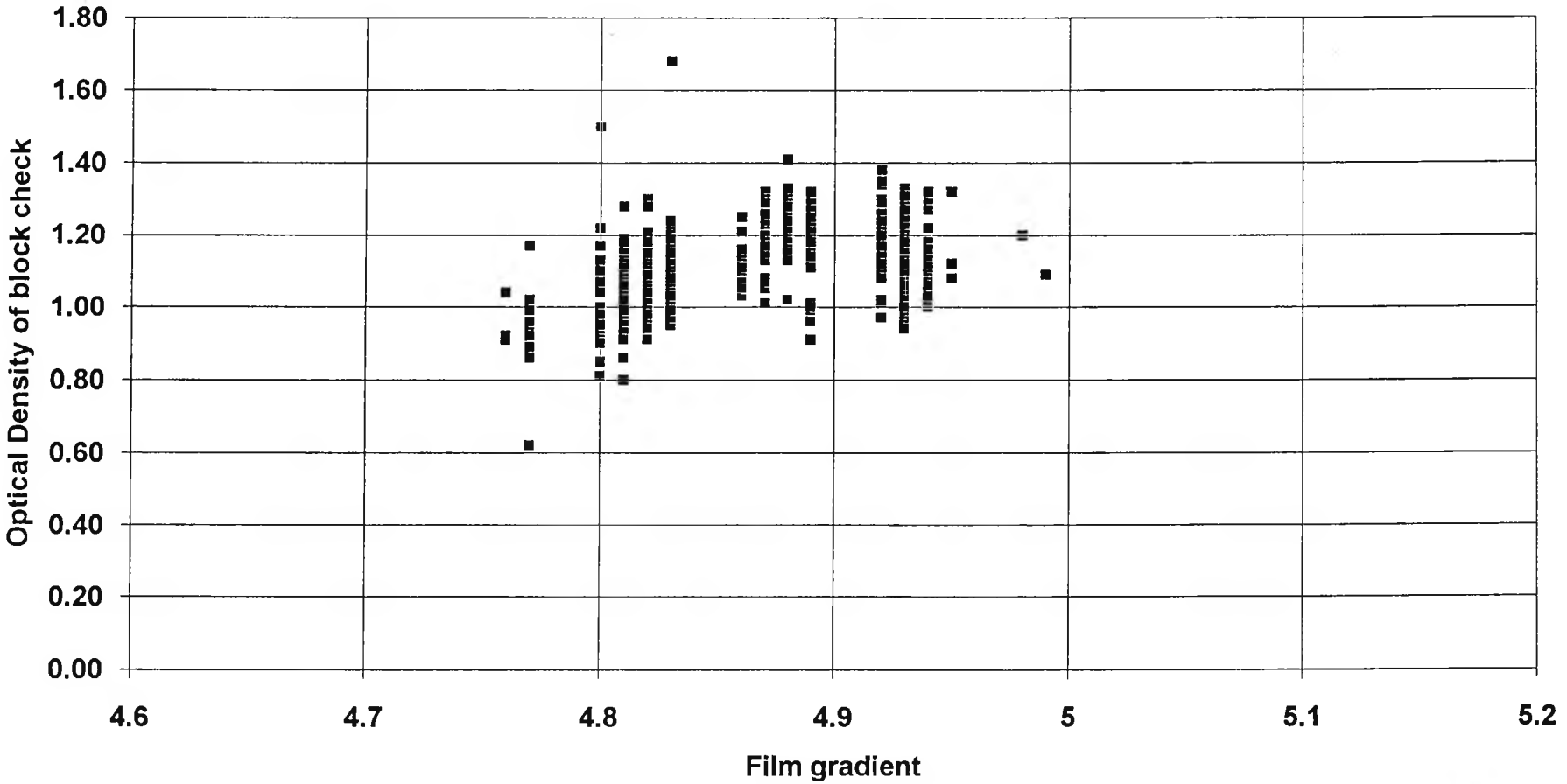
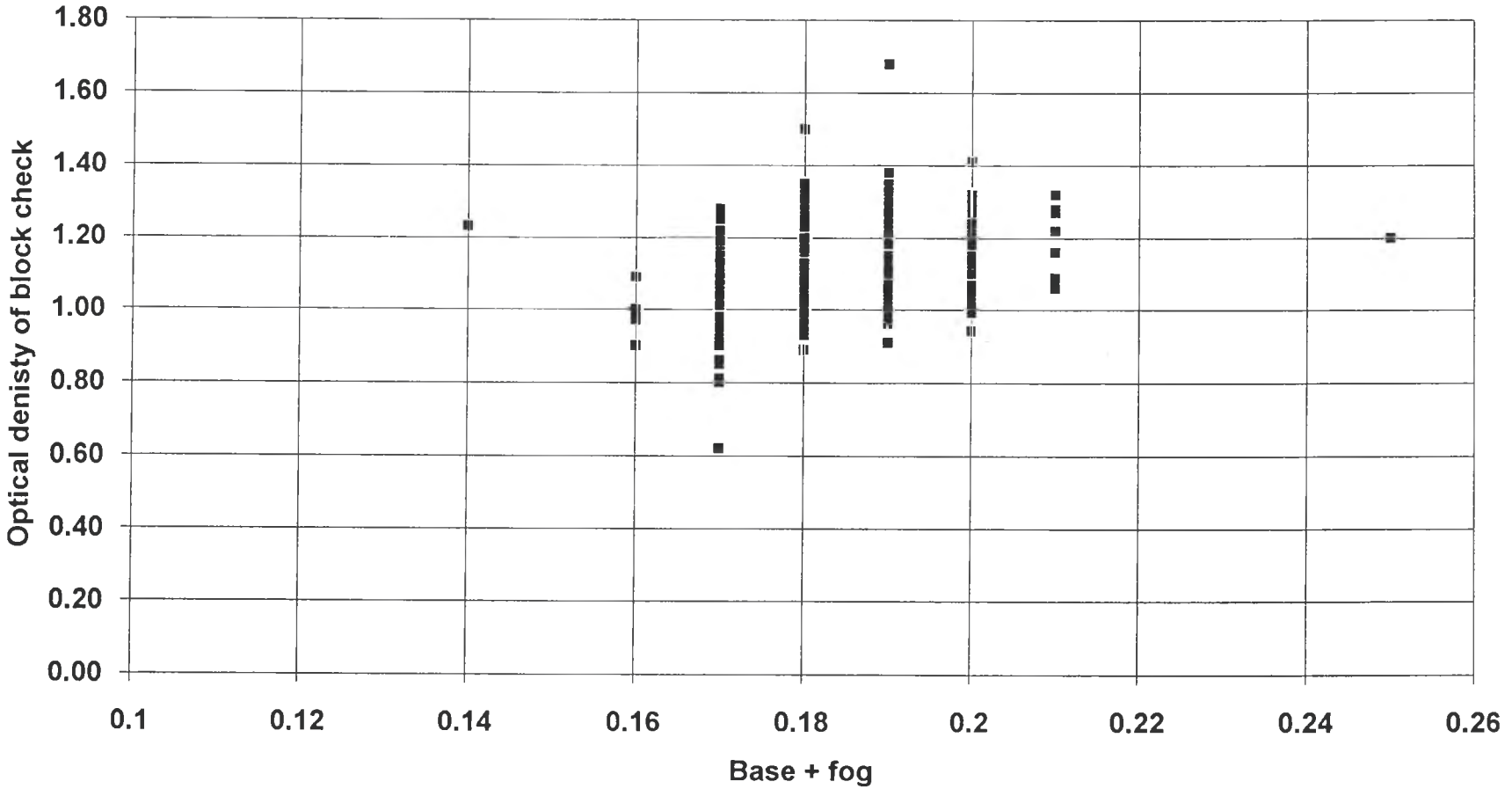


Figure 4.2.1.4

Relationship between fog and daily block check density at Whipps Cross



Date	mAs	Image Score	Fog	Speed	Gradient	Dmax
9.10.90	39.2	9	0.17	292	4.87	3.61
11.2.91	41.2	4.5	0.20	293	4.88	3.77
1.3.91	40.1	7	0.18	297	4.82	3.78
11.6.91	40.1	5	0.2	317	4.81	3.90
9.6.92	40.7	6	0.17	364	4.77	4.19
2.2.93	40.5	5	0.19	362	4.87	4.57
25.5.93	44.4	6	0.18	319	4.92	4.55

Table 4.2.2 - Parameters describing the characteristic curve and corresponding image quality scores from physics visits

The correlation coefficient has been calculated for each of the graphs in figures 4.2.2.1 to 4.2.2.4. The only parameter which shows a statistically significant correlation with image quality is the fog level, for which the correlation coefficient is -0.784 ($0.02<p<0.05$). For the same data, a multiple regression has been performed, the results are shown in Appendix E.

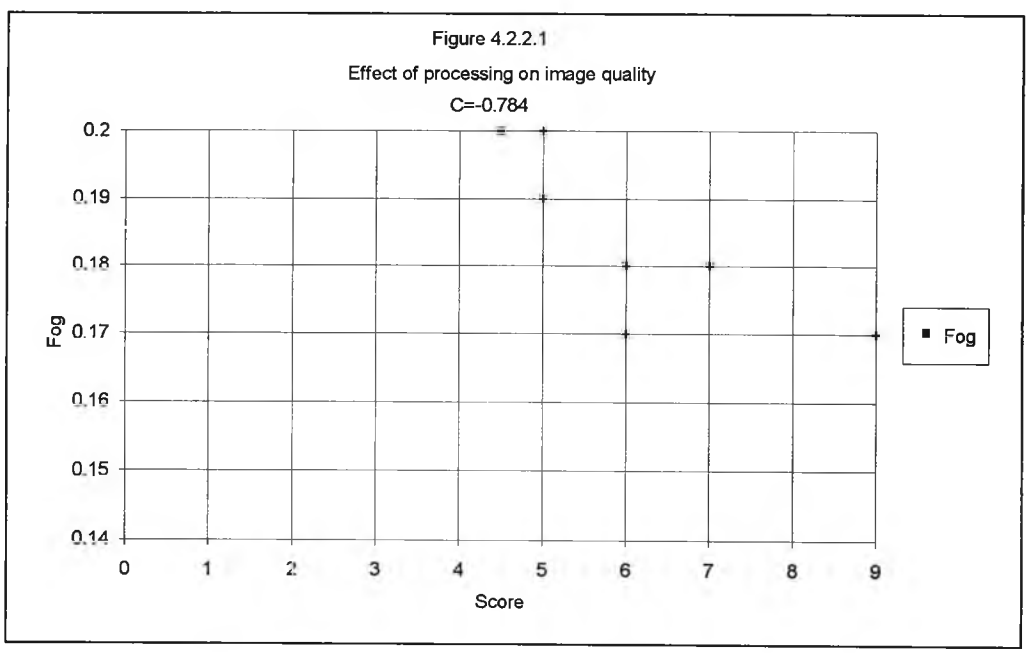


Figure 4.2.2.2

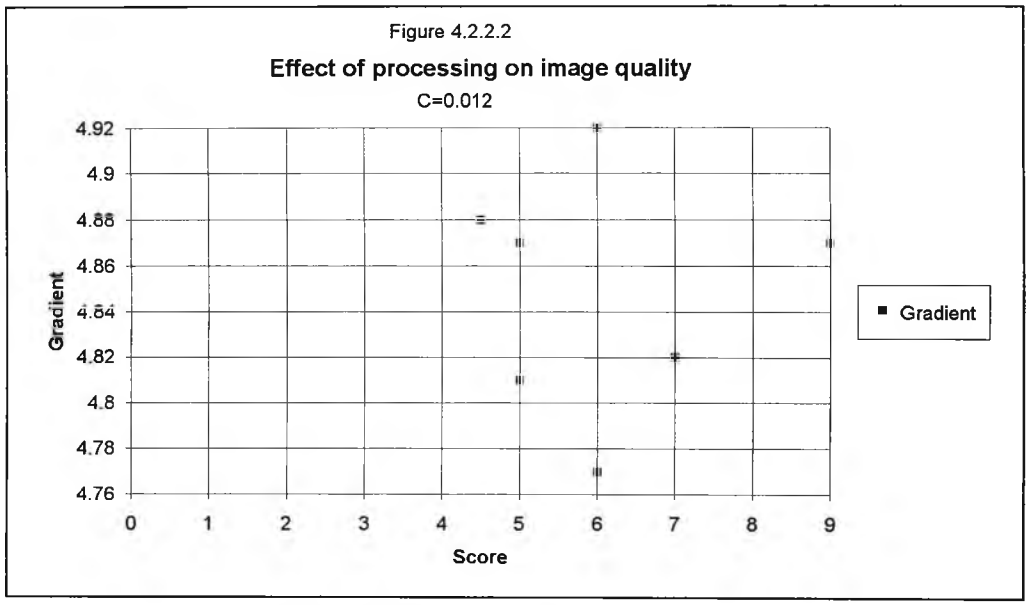


Figure 4.2.2.3

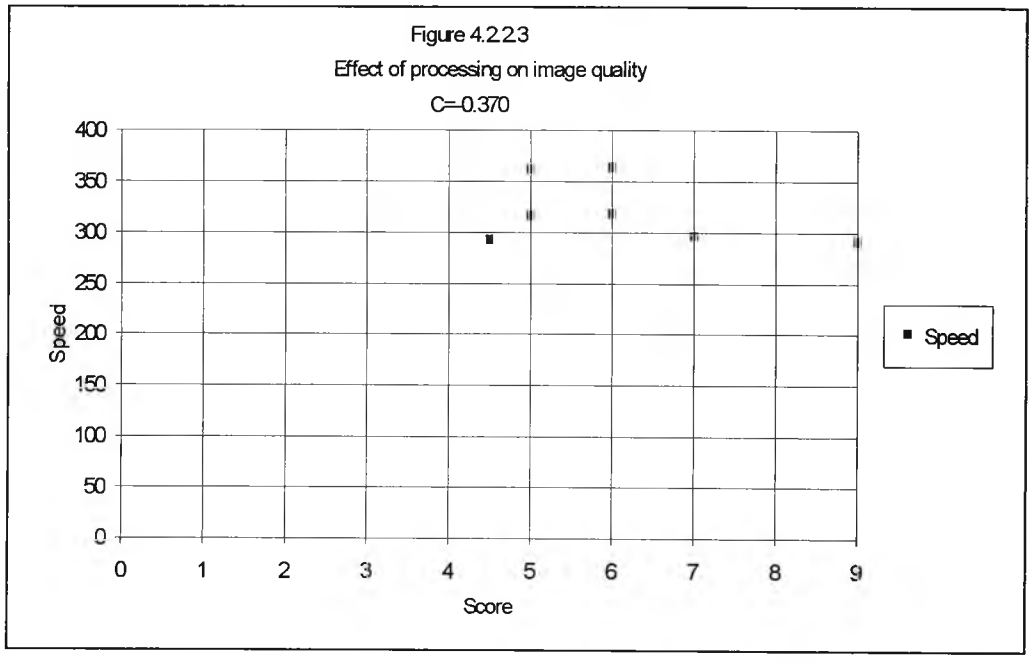
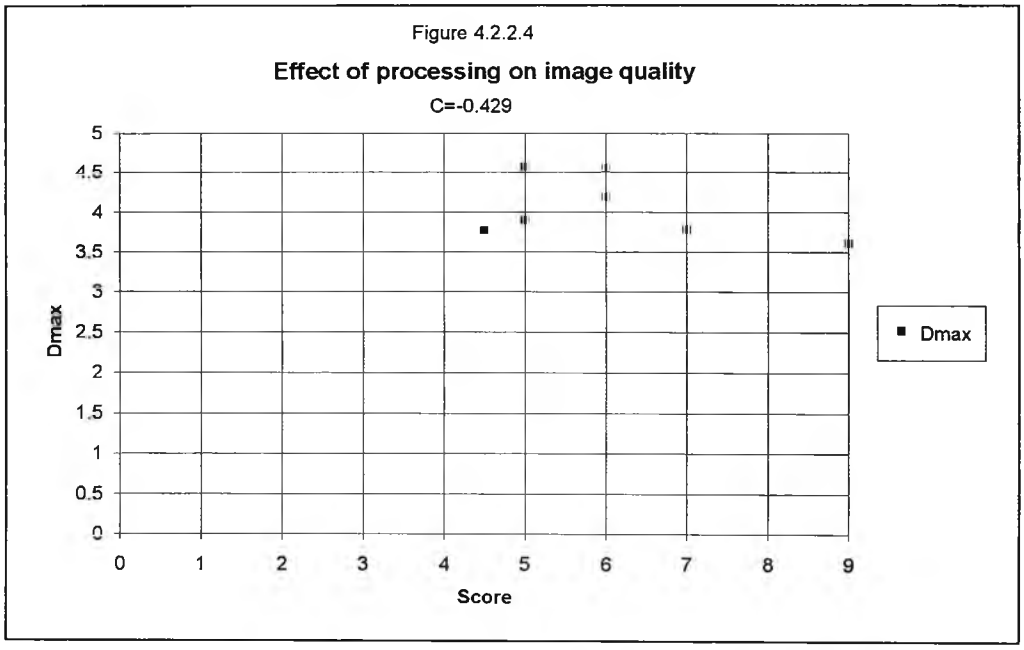


Figure 4.2.2.4


Effect of processing on image quality
C=-0.429
Dmax
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
0
1
2
3
4
5
6
7
8
9
Score
Dmax

Chapter 5

The Imaging Chain

This chapter considers how the information gathered and the knowledge of the relationships between the various measured parameters can be used to produce a more cost effective way of doing quality assurance. A proposed scheme of work is produced and financial comparisons made between the old system under which the data was acquired and the new system which has recently been implemented.

5.1 Introduction

Sensitivity analysis of the breast screening system showed that a major concern with breast screening is that good image quality should be established and maintained otherwise the sensitivity, specificity and predictive values of the screening process suffer. Good image quality in measurable terms is a combination of spatial resolution sufficient to see the smallest object of diagnostic significance and the ability to reproduce changes in X-ray intensity which correspond to tissue structures, or contrast resolution. This is always constrained by the need to keep the radiation dose to a level where harm to the screened population is minimised. In practical terms, this means that during the commissioning process, the imaging system should be optimised and, thereafter, each contribution towards the final image should be monitored for consistency. If the characteristics of the individual components of the imaging system (the tube, AEC etc.) have not changed, the image quality must be the same. What is required of a monitoring system is that it should be able to tell us whether any of the component characteristics have changed and consequently whether the image quality is as good as when the system was first commissioned.

If a system as a whole does go outside of its control limits, it is necessary to identify which part of the imaging chain is responsible, in order to take remedial action. The monitoring system should be capable of doing this if the measured quantities are carefully selected and the relationships between them are well understood. The natural variation of each part of the system should be known so that any excursions observed are statistically valid and time is not wasted looking for a fault where there is none.

What follows is theoretical and proposes a system which could be implemented on a computer and used by radiographers with the minimum of training. The philosophy is to identify a set of measurements which are able to be read and analysed by a computer in order to eliminate, for example, random variations in image quality score due to the scorer. These measurements should be capable of identifying which part of the system is at fault, thus enabling rapid further checks to pin down the fault. The second aim is to be cost effective in performing quality assurance.

The imaging process is more fully described in chapter 2, the following summary identifies the key features required for the discussion to follow.

5.1.1 X-ray production

A heated filament, the electron gun [a], produces free electrons which are accelerated through a vacuum by a high voltage [b] to the anode [c]. The electrons interact with the target material to produce X-rays [d]. These pass first through the beryllium window of the X-ray tube [e], then through any added filtration [f] which is designed to alter the X-ray spectrum in a controlled manner.

5.1.2 Image formation

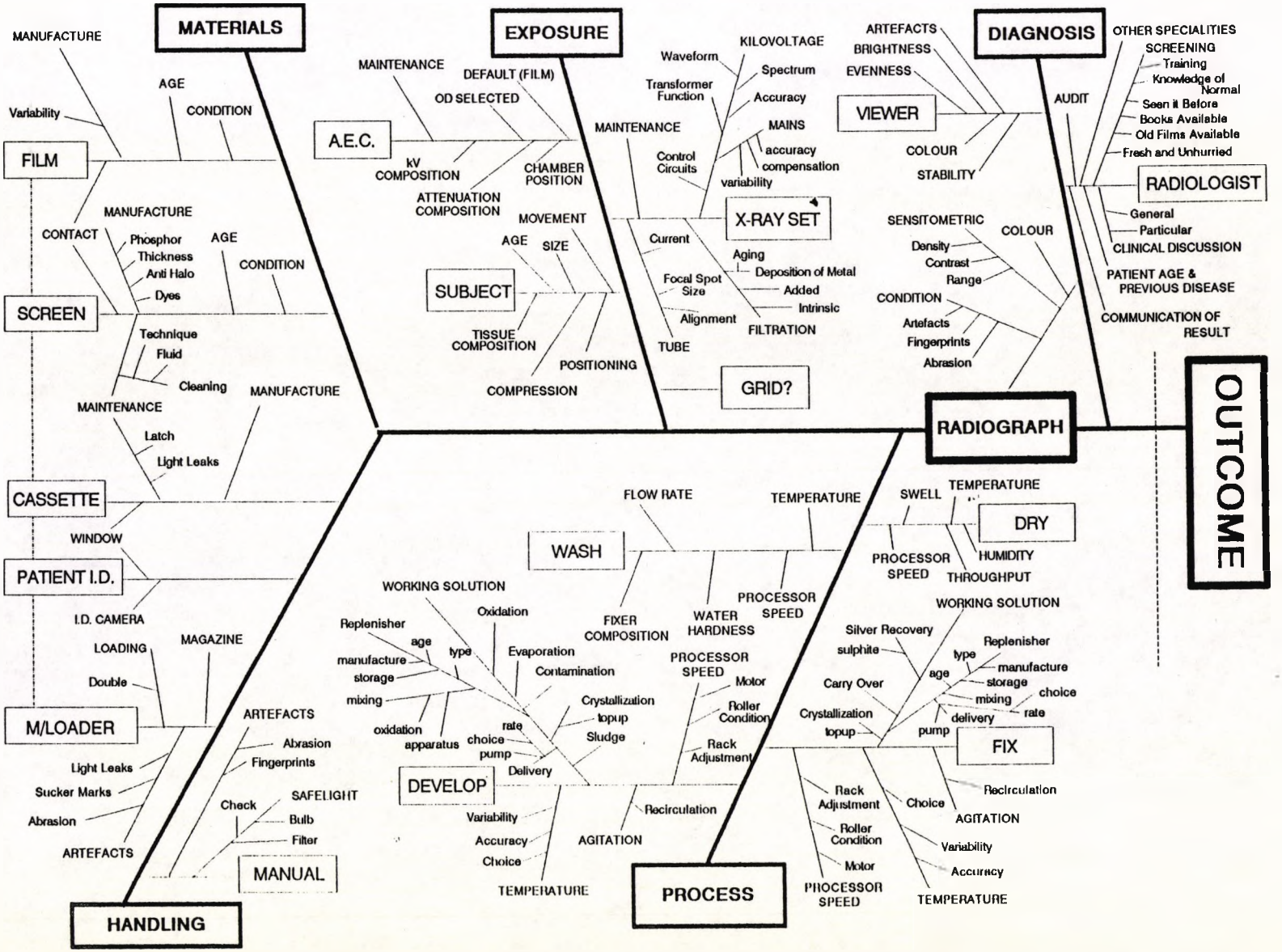
The image formation part of the process occurs when the X-rays are differentially absorbed as they pass through the breast. This part of the process is not reproducible, as each woman is different, but it can be mimicked by using reproducible attenuating materials [g] . The radiation then passes through an anti-scatter grid [h] which removes most of the scattered radiation, which contributes only to the noise, and a proportion of the primary beam which would have contributed to the image had the grid not been there. After passing through the front of the cassette [i] the radiation strikes the intensifying screen [j] where it is converted to light; the light from the screen interacts with the emulsion on the films [k] to form a latent image. Any radiation which does not interact with the screen passes through the back of the cassette and hits the automatic exposure device [l] which is designed to terminate the exposure once the appropriate amount of radiation has been received.

5.1.3 Processing

When the film is fed into the developer tank [m], active sites within the emulsion where light had interacted with the film are "developed" into silver-containing grains. The endpoint of this process depends on the temperature of the chemicals [n], their concentration [o], and the time spent in the developer [p]. The film passes through rollers, which squeeze out excess developer and into the fixer tank [q]. This process serves to prevent the reaction from proceeding any further. The film is then washed [r] and dried [s].

All of the factors which have an influence on the final image have been placed on an Ishikawa [146] (fishbone) diagram in figure 5.1.3. It is impractical and undesirable to monitor every single parameter listed; for the purposes of quality assurance groups of factors may be monitored by a single test, for example, all of the items associated with processing [m,n,o,p,q,r and s] can be monitored as a group by using sensitometry. As long as the sensitometry parameters are constant, it can be assumed that all the factors which contribute to that outcome are also constant. Should there be a measured change, two different methods of analysis may be used; first, the results can be combined with other results to infer the cause of the change, second, further testing may be required in order to discriminate between the various parameters which contribute to the gross change. One such example might be a speed increase during sensitometric testing which would lead to measurement of temperature and chemical concentrations in order to find the cause of the change. Either or both of these methods may be required to produce the required diagnostic information.

Figure 5.1.3 Ishikawa diagram of the diagnostic decision in mammography



5.2 Record Keeping.

In order to make data gathering as efficient as possible and avoid the collection of redundant data, a minimum data set which will enable the contribution of all stages of the imaging chain to be specified needs to be identified. For each parameter affecting the image quality, some possible tests are suggested

a electron gun

If the electron gun starts to deteriorate, there are two possible effects, the focal spot size may change as the filament starts to warp and the output of the tube may change if any filament material has evaporated. The output may be altered either by change of filament resistance or deposition of filament material on the exit window both of which would be due to overheating of the filament. If the filament has become narrower because of tungsten evaporation, its resistance will have increased, this means that a smaller current will flow and therefore the filament will be heated to a lower temperature and the tube current will drop, causing a drop in exposure rate at any set kV. Any evaporated tungsten which may have been deposited on the window of the X-ray tube will cause hardening of the beam and will decrease the output of the tube.

To test for a change in filament size due to thermal overheating, the resolution or the focal spot size should be measured. The resolution can be determined using a bar pattern. The focal spot size can be measured by several methods: slit camera, star pattern or pinhole.

Reduction in tube current manifests itself in longer exposure times. As a consequence the automatic exposure control (AEC) may have problems in compensating for reciprocity failure, resulting in radiographs which are under-exposed. Exposure time can be measured directly or inferred from the post exposure mAs.

Deposition of material on the window will decrease the output per mAs, thus increasing the mAs recorded under AEC and also hardening the beam which may alter the contrast of the final image. This could also be evaluated by measurement of the HVL, the output or a contrast index.

b kV

The voltage across the tube is controlled electronically or electrically external to the tube. It is possible to have either a long term drift or temporary changes which return rapidly to the equilibrium value. Changes in kV will affect both the output and the beam spectrum, causing a change in recorded mAs under AEC and a change in contrast, respectively. These effects can be caused by other fault conditions, therefore the kV needs to be separately measured to establish the cause of the post exposure mAs changes or contrast changes.

Two suitable direct methods are:
a) using a kV meter or
b) invasive measurement with an oscilloscope at the test point;
indirect methods are step wedge measurements and monitoring the post exposure mAs for blocks exposed under AEC.

c anode

The condition of the rotating anode can affect the image quality in two ways; if it has worn bearings and consequently wobbles, it will cause the effective focal spot size to increase; if there are cracks present or pitting on the surface, the distribution of electrons on the anode surface will change causing a change in X-ray distribution. The former will affect the resolution which can be monitored in the ways listed for the electron gun, the latter will alter the output spectrum which can be monitored using blocks exposed under AEC, air kerma measurements, HVL or tube current measurements. The anode material also affects the X-ray spectrum, but this does not change over time, and as the anode is in a vacuum, the anode does not oxidise. To test for warping of the target due to thermal overheating, the resolution or the focal spot size should be measured.

d X-rays

The quality and quantity of the X-rays are a function of the X-ray tube kV, the anode material and the tube current. Each of these parameters which change the nature of the X-rays being produced are considered separately. It is not possible to measure the X-rays before they have been

filtered as this would entail inserting a sensor into the vacuum tube which would destroy the vacuum. The X-rays are completely dependant on tube current and tube voltage, variables which are already being measured.

e window

The filtration due to the window may change due to deposition of vaporised tungsten from the filament or vaporised molybdenum from the anode. Suitable tests for monitoring the window filtration are HVL measurements, contrast measurements and AEC measurements.

f filter

Mechanisms by which the additional filtration might change are mechanical damage or oxidation of the filter material. These cause the output to be altered and also change the spectrum of the beam. Suitable measurement methods are X-ray step wedge exposures, blocks exposed under AEC and HVL measurements.

g attenuating materials

In normal use, the attenuating material is the breast, and the variations seen are normal and desirable; they are, in fact, what the imaging process attempts to record. During testing, the attenuating material is normally perspex. This can vary slightly from batch to batch and the thickness of each piece of perpsex is not always the same as its nominal value. However, if the same perspex is used each time, then there will be no variation from this source, this has been incorporated into the protocol for measurement.

h grid

The grid would not be expected to alter unless it had undergone some mechanical damage. This would be evident as an artefact on images and can be best monitored by visual inspection of plain films. There could also be problems if the mechanism by which the grid is moved has ceased to work correctly. This could cause grid lines to appear on radiographs and is again easily monitored by visual inspection of radiographs.

i cassette

The attenuation of the front face is the only part of the cassette which affects the radiation reaching the screen and film, but, because the AEC is positioned behind the cassette, the overall attenuation of the cassette is important since it alters the overall exposure. In order to avoid such variations, the protocol demands that the same cassette should be used in testing. The block checks would indicate any decrease in the cassette attenuation, perhaps by abrasion, by a decrease in mAs with a constant density, however further tests would be required to determine whether this or some other change had caused the drop in mAs.

j screen

The sensitivity of the screen and the attenuation of the screen, both of which are functions of the screen thickness, may vary due to abrasion, and the sensitivity may diminish due to chemical changes within the screen. This may be monitored using a combination of sensitometry tests, step-wedge measurements and block checks.

k film

The characteristic curve of each piece of film can be determined using light sensitometry. This, however, includes the effect of the processor. It is expected that there will be small random fluctuations from film to film and somewhat larger fluctuations from batch to batch. It is also important to note that this test is not the same as an X-ray stepwedge. The response of the films depends on the spectrum of the light source which in general is not the same for a sensitometer and an intensifying screen.

l AEC

Block checks are used to determine the function of the AEC. Its job is simply to terminate the exposure when the appropriate amount of radiation has reached the film. The AEC has some built in compensation for dose rate and beam quality, but this is not perfect and a drift in kV or output may cause the apparent function of the AEC to drift. The cause of

an mAs drift can be ascertained by considering other tests in combination with the block checks. To determine the cause of an optical density drift, the film sensitometry needs to be considered in conjunction with the block checks.

m developer tank transport mechanism

Problems within the developer tank which are not due to temperature or chemical concentration errors may be determined by visual inspection of a plain film for pick off, roller marks and other artefacts.

n temperature of developer

The temperature of the developer is thermostatically controlled and larger units have a temperature display, however, these may not be correctly calibrated and thus show a consistent but incorrect temperature, or a fault can develop so that they no longer function correctly. Any temperature fluctuations due to such faults are expected to show up as a change in sensitometry. Positive diagnosis of a temperature fault can be made using a thermometer which is independent of the temperature display, electronic thermometers were issued for this purpose.

o developer concentration

Problems within the developer concentration should show up in the sensitometry results. Specific gravity measurements, chemical analysis and measurement of replenishment rates can be used to differentiate this type of fault.

p development time

Over-long or over-short development times caused by variations in the speed of the motor driving the transport mechanism will affect the sensitometry. A long development time causes increased optical density and decreased contrast due to loss of selectivity. The overall processing time can be measured with a stopwatch in order to establish if the development time, which is always a fixed fraction of the processing time, is correct.

q fixer tank

Problems within the fixer should show up in the sensitometry results, although visual inspection is often sufficient to diagnose fixer problems, with milky streaking becoming apparent.

r wash tank

Sticky films or films which smell strongly of the fixer chemicals indicate that the films are not being properly washed. Over washing is not a problem if the film has been correctly developed and fixed.

s drier

A simple visual inspection is sufficient to see if the films are properly dried.

There is generally more than one technique for evaluating the performance of any particular component of the system; in order to minimise cost one should take advantage of the overlap so that two measurements are not made when one would suffice. With the exception of the focal spot and/or resolution measurements, all of the contributing parts of the system can be monitored by three sets of measurements: light sensitometry to monitor the processor, X-ray sensitometry (stepwedge measurements), and exposures of blocks of perspex using the post exposure mAs meter as an uncalibrated dose-measuring instrument. There would still be a need for periodic measurements of absolute values to recalibrate the system. Table 5.2.1 shows the relationship between selected tests and the components of the imaging system which affect their outcome.

Fault	PC Sensitometry	Block mAs	Block OD	Block Contrast Index	X-ray Stepwedge (processor removed)	X-ray stepwedge (raw)
Processor	Y	N	Y	Y	N	Y
kV	N	Y	?	Y	Y	Y
Output intensity	N	Y	N	N	N	N
Filtration	N	Y	?	Y	Y	Y
Focal Spot	N	N	N	N	N	N
AEC		Y	Y	?	N	N
Screen Sensitivity	N	?	Y	N	?	Y
Film	N	N	Y	Y	N	Y

Table 5.2.1 The relationship between tests and potential source of faults. "Y" in a box indicates that the fault in that row causes a change in the test in that column, "N" that the fault does not affect the test and "?" that the fault may affect the test depending on how the whole system is working. Such a test could not be used as a diagnostic tool but may be useful for differentiating between fault conditions.

5.3 Diagnostics.

It has already been stated that monitoring of every item contributing to the imaging process is not necessary, desirable or possible. Instead, a selection of tests which are simple to perform need to be carried out regularly, these tests need to be able to monitor *all* of the imaging chain and to provide enough information to diagnose which particular component of the imaging chain is causing the gross effect. This section is an outline of the methods necessary to separate all of the contributions to the overall variations. These ideas can be used to integrate the efforts of the physicist with those of the radiographers and to tie in output measurements and kV, which are done infrequently, with mAs measurements, which are done daily; the relative contribution of cassette variations, film variations, process variations, X-ray spectrum variations and automatic exposure control to the final image can also be estimated.

5.3.1 Block checks

A block of perspex is placed on the breast table where the breast would go and, with a loaded cassette in the bucky, an exposure is made under AEC; the post exposure mAs is recorded, the film is developed and the density of the film at a fixed point is measured. This test is sensitive to changes in output which may be due to filtration, kV or tube current changes, and also monitors the AEC function. The density of the developed film is useful in differentiating between faults due to the AEC which alter the quantity of radiation reaching the cassette and consequently alter the density in line with the change in mAs, and those which alter the spectrum of the radiation causing a change in mAs but a fairly constant optical density because the AEC has successfully compensated for changes in the X-ray system. If the filtration has changed, the AEC cannot detect that this is not simply a woman whose breast has a particularly high or particularly low attenuation coefficient and attempts to produce an exposure which will result in the target OD being achieved.

5.3.2 X-ray sensitometry (stepwedge measurements)

In this test a perspex stepwedge is placed on the breast table, a block of lead is placed along the film edge furthest from the chest wall and an

exposure is made under manual control. This ensures that the value of the mAs is the same for all measurements and its value will have been determined in advance to produce an exposure which will give a good representation of the X-ray sensitometric curve. The area of the film which has been protected by the lead block is then exposed to a light sensitometer. This enables the effect of the processor to be removed and a conversion made to radiation exposure (in arbitrary units). The fact that both X-ray and light stepwedges were produced on the same film reduces the variation which can be attributed to the film or the processing, it also reduces the amount of raw material required for QA tests. The X-ray stepwedge test is therefore sensitive to both the processor function and the image formation process whereas the light sensitometry is sensitive to the processor function alone. The relationship between optical density and light exposure, the equation of the characteristic curve [147], is given by

$$D = \frac{D_{saturation}}{\left(1+\left[\frac{\tilde{E}}{E}\right]^{\beta}\frac{1}{\alpha}\right)^{\alpha}} \qquad \text{Equation 5.1}$$

where:

$\tilde{E}$ is the exposure at the point of inflection of the curve

E is the exposure producing an optical density D

β is related to the gradient of the curve

α is the degree of asymmetry of the curve

$D_{saturation}$ is the maximum optical density which the film is able to produce

Using this equation with the sensitometric wedge, we can ascertain exactly how much exposure is required to produce any given optical density; the film is acting as a dosemeter and the light sensitometry is used to calibrate it. The stepwedge can then be used as a kV meter. The ratios of the exposures under each step of the wedge define the attenuation coefficient of the wedge material, perspex (poly methyl methacrylate); the attenuation coefficient is kV dependant. Any variations in relative exposures of steps must be due to kV variations because all steps were exposed simultaneously. The limitations of the technique are:

1 There are variations from left to right in radiation intensity due to the geometry of the system. From knowledge of the X-ray unit geometry, a calculation of the effect can be made and removed from the raw data.

2 There could be incongruities due to non-homogeneity of the filtration or misalignment of the tube. The blocks are uniform, therefore any non-uniformity which is observed must be an integral part of the system; the geometry of a mammography unit is fixed and is the same for the step wedge as it is for the perspex block, therefore, a correction for non-homogeneity of the beam can be made by reference to the film produced during the block checks.

3 Ratios rather than absolute values should be used to eliminate any effect due to variations in the radiation output from one exposure to the next. The corollary is that if absolute values are used for one particular step, the test object can be considered to be acting as a dose meter. It would be interesting to investigate how well this corresponds to measurements of air kerma made in the conventional way.

Contrast measurements at kV from 25 to 30 indicate that variations of μ with kV are very small and difficult to detect.

5.3.3 Process Control Sensitometry

A light sensitometer is used to create a series of steps each of which varies from its neighbour by a logarithmic intensity factor of 0.15. A box of films is put to one side and kept specifically for the purpose of monitoring the state of the processor. This means that for 100 working days (approximately 20 weeks) any changes to the result which are caused solely by the processor (this assumes that there are no changes in the output of the sensitometer). When a box of film has nearly all been used, a second box of film must be opened and for a period of four or five days, a strip is printed on films from both boxes which are processed in quick succession. If the two films do not produce identical curves, corrections can be made to later sensitometry strips or their control parameters. If the

two boxes are from the same batch, one would not expect there to be any statistically significant variations from one box to the other.

The processor is the final link in the imaging chain and its function affects any test which uses film. It is therefore vital to remove the effects of any processor variations from any film-based measurements. By using the sensitometry curve, these changes can be related back to the amount of light which was required to produce that optical density in the first place, this then enables a correction to be made to any test film to take the density back to the reference situation. Software to do this is being developed by Mr A Green of Kodak Research but has not yet been released.

It is also important to monitor the state of the processor routinely in order to ensure that the processor will give satisfactory results in normal use. It is to be expected that there will be fluctuations from day to day, and we wish to ensure that these variations are within statistical control limits. For a measuring system however, it is necessary to correct for these normal fluctuations.

5.4 Fault Detection

The selected tests have been chosen for their ability to monitor the whole of the imaging chain, thus if all of the test results remain within their specified limits, the system as a whole must also be within its limits. When one or more of these tests give results which indicate a fault condition, it is necessary to identify exactly which part of the chain is responsible for the fault. In order to do this, the relationships between the test parameters need to be established, making it possible to identify ways in which each element of the imaging chain can be monitored by manipulating the data set which has been generated.

5.4.1 Processor

Monitoring the processor is quite straightforward and has been fully described in section 5.3.3. The usual rules of statistical quality control are applied and remedial action taken when a trend is detected or when a specified limit, which would usually be three standard deviations, has been exceeded.

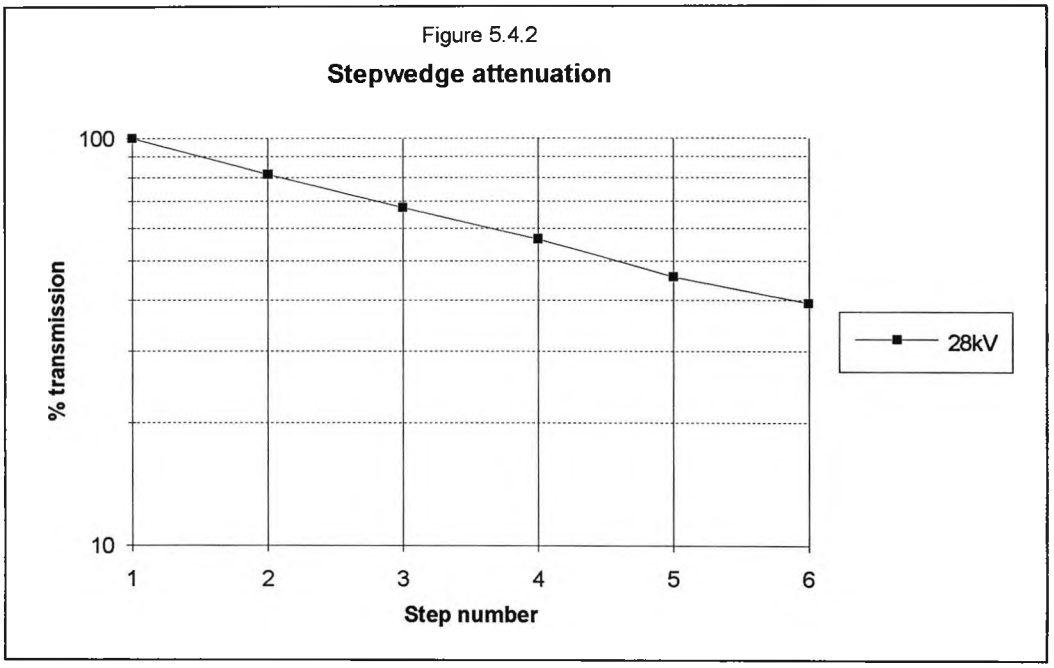
This monitoring serves two purposes; firstly, to enable corrections for processor variations to be made to all of the other parameters which are evaluated using film, secondly, and more importantly, it prevents the processor from being used in a sub-optimal condition.

At Whipp's Cross Hospital, this was illustrated when the water filter which feeds into the automixer became blocked. The water flow was reduced, the replenishing chemicals became more concentrated and the sensitometry was affected. The detected change warned that investigation was required and the problem was quickly remedied. If the development process is well understood by the film and/or chemical manufacturers, it is possible to provide a computerised diagnostic feature which is able to suggest possible causes for the variations in the processor based on the sensitometric curves in its memory.

5.4.2 kV

Initially, the system would need to be calibrated by experimental measurements, which could be updated on the computer using measurements made every six months by the physicist with calibrated equipment. This would enable fundamental measurements, such as kV, to be extracted from the composite measurements. Using kV measurement as an example, if a kV meter was not available (and unless there is no alternative way of measuring the kV, it would be uneconomic to have a kV meter, costing about £1900 at 1993 prices, for every X-ray unit), then it would be possible to monitor kV by doing daily exposures of a perspex stepwedge which have a sensitometric strip printed on the opposite edge. The stepwedge exposure would be under manual control to eliminate the performance of the AEC as a variable. First, the light sensitometry strip is read by the computer and equation 5.1 is fitted to the data. We then know how much radiation was incident on the film for any given optical density. (The true situation is somewhat more complex than this because the direct X-ray exposure contributes between 10% and 20% of the exposure, the vast majority of the exposure comes from the fluorescent screen in the form of visible light.) It is possible to plot the attenuation of radiation through the stepwedge (Figure 5.4.2) on a semi-log plot it forms a straight line (neglecting the effects of beam hardening).

Figure 5.4.2



This illustrates Beer's law, that radiation passing through a homogeneous medium is attenuated exponentially

$$I = I_0 \exp(-\mu t) \qquad \text{Equation 5.2}$$

where:

μ is the attenuation coefficient of that material

t is the thickness of the material

I is the intensity of radiation at a depth t

$\mathbf{I_0}$ is the intensity of the radiation before is begins to pass through the attenuating material

This however is only true for a single energy of radiation because μ is energy dependant. So, for an X-ray spectrum with a range of photon energies, because μ is greater for lower energy radiation, this is preferentially absorbed and the spectrum is altered (hardened) as it passes through an attenuator. Thus when we measure the intensity of the radiation causing the blackening on the film the ratio of intensity from one step to another depends on the spectrum as well as the attenuation. Consequently, in order to get accurate results, the system has to be calibrated. The calibration will be unique to the X-ray unit, although it will be similar for similar types of system.

5.4.3 Output of tube per mAs (exposure rate)

If the output per mAs, or specific output, of the tube falls, the mAs required to trigger the automatic exposure termination will rise. If these changes are sufficiently small that there is no reciprocity failure, the optical density on the film will not change because the total amount of radiation reaching the film is unchanged.

It is not possible to predict whether contrast measurements and the X-ray stepwedge results will change because they depend on the mechanism by which the specific output has changed.

5.4.4 Filtration

If the filtration changes then both the radiation spectrum and the output per mAs will change. The mechanisms by which this might happen are

(i) deposition of the filament material (tungsten) or anode material (molybdenum) on the beryllium window as a result of excessively high temperatures

or

(ii) corrosion of, damage to or removal of the external filter.

In both cases, the radiation spectrum will change as well as the specific output. Monitoring the output per mAs or the post exposure mAs as an indirect measure of this quantity will enable the filtration to be monitored. If such a change has been detected, the source of the change must be found. Visual inspection of the external filter allows its condition to be assessed. Internally the output per mAs may have dropped because of deposition on the window or by pitting of the anode, which causes a small scale version of the anode heel effect, thus hardening the beam.

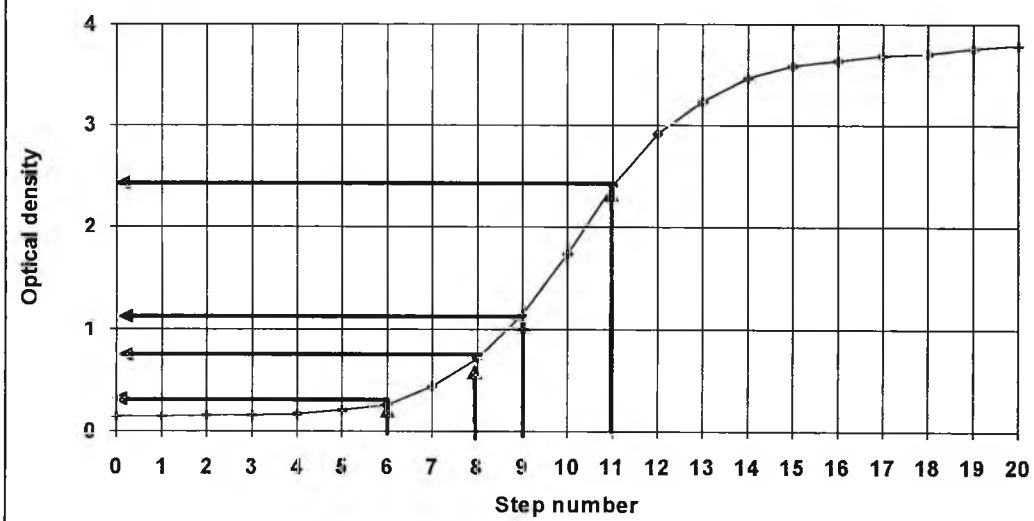
5.4.5 Automatic Exposure Control

If the AEC malfunctions, the characteristic feature will be variations in the post-exposure mAs which bear no relationship to the mAs required for a uniform density. Therefore, these mAs changes would be accompanied by

changes in optical density which have a strong correlation with them. There will not be an exact correlation because the processor also has an effect on the final density produced, but since we measure the state of the processor with a light stepwedge on the same piece of film, the variation of the processor can be mathematically removed and the correlation would be good. The stepwedge does not depend upon the AEC and would therefore be unaffected.

The block contrast index, the difference in OD between a high attenuation and a low attenuation area, will change as the gradient of the operating point changes as illustrated in figure 5.4.5.

Figure 5.4.5 The relationship between contrast index and characteristic curve



5.4.6 Screen Sensitivity

It is not yet known whether the sensitivity of intensifying screens varies over a period of time, but it has been demonstrated that there can be considerable variations between nominally identical screens. Acceptance tests at one breast screening unit showed that the optical densities produced under AEC ranged from 1.01 to 1.21 with a mean of 1.09 and a standard deviation of 0.08 giving a coefficient of variation of 7.0% (n=12). The optical densities corresponded to changes in mAs and were most probably due to variations in the thickness of screen material. However, the magnitude of the change in optical density was much greater than that

in mAs, for which the coefficient of variation was 1.65%. This results from the fact that a greater thickness of screen material interacts with a greater proportion of the X-rays and causes two effects simultaneously, more light is produced per mAs because a greater proportion of the available radiation is absorbed, and more radiation is required (greater mAs) to terminate the exposure because there is less radiation transmitted through the thicker screen. Combining these, there is an increase in optical density resulting from the combination of both effects when the screen is thicker.

It is reasonable to suppose that regular cleaning of the intensifying screens gradually abrades material from the screen surface and reduces the apparent speed of the screens in normal use, when automatic exposure control would be used. It is also possible that the phosphor deteriorates chemically over a period of time. Consequently it is wise to monitor the screen speed over a period of time.

The screen of the test cassette will be monitored automatically as a result of the routine measurements. The screen speed which can be monitored by looking at the speed from the X-ray stepwedge once the effect of film and processor variations have been removed (the X-ray stepwedge is normalised to day 1 of testing). The other cassettes would need to be periodically tested. All of the required information required can be obtained from two films: the film of the X-ray stepwedge with light stepwedge to monitor the speed of the screen to a fixed amount of radiation, and the block test to monitor the attenuation of the beam by the screen.

The screen only interacts with the system at the stage where the X-rays hit the screen. If the screen changes are caused by mechanical abrasions, the attenuation of the beam by the screen will be reduced and consequently the mAs under AEC will be decreased. There will be a corresponding reduction in the optical density. On the stepwedge however, the exposure is manually controlled and therefore the curve ought to be the same.

If the deterioration of the screen is caused by chemical breakdown, the mAs is unlikely to change but the optical density of the film certainly will. The contrast index will probably change as will the X-ray stepwedge.

5.4.7 Film Variations

Any batch to batch variation of the film will show up in the following tests:

(i)	4 cm perspex block	-	density change
(ii)	contrast block	-	density and/or contrast change
(iii)	X-ray stepwedge	-	speed and/or gradient

The sensitometry test uses film from a box which has been set aside in order to eliminate batch to batch variations, the light generated characteristic curve will therefore not change. So if the processor has been shown not to vary, or the effect of its variation has been mathematically removed, the light stepwedge on the back of the X-ray stepwedge film can only have changed because of the film batch.

Measurements made in NE Thames on Kodak films show that variations from batch to batch are smaller than those due to the processor in routine use. The batch to batch variations have no practical significance and one only needs to take account of them in experimental work where high precision is required. Similar measurements on DuPont films show variations which are significant when measured by light sensitometry but, when X-rays are used to generate result, the variations become very small. This is thought to be due to the light spectrum of the sensitometer being much broader than the spectrum produced by the intensifying screen so that small differences in speed are exaggerated when measured by light sensitometry.

5.4.8 Focal Spot

Variations in focal spot size cannot be detected by the four itemised tests. The focal spot size tends to affect the sharpness of the image but little else.

Deterioration of the focal spot can be due either to deterioration of the filament, which will generally be accompanied by a deposition of filament material on the walls of the tube, or by damage to the anode.

Tests which can be used to measure the size and condition of the focal spot are:

(i) pinhole camera
(ii) slit camera
(iii) star pattern
(iv) resolution grid (e.g. in a test object)

The first three tests are time consuming and would not be performed by the radiographer: they would form part of the less frequent but more comprehensive physics tests. The resolution test could be included in an area of the contrast test object. As things stand, this would have to be quantified manually although, as technology develops, it should be possible to produce a resolution test which is machine readable.

5.5 Financial Analysis of New Scheme

Under the old system, the physicist was scheduled to visit every unit three times per year. One of the three visits was an annual visit where the checks would be more comprehensive than the routine checks. The radiographers tested the AEC with a 4cm block of perspex on a daily basis and also tested the processor function using light sensitometry.

In addition, when the mobiles moved, they were given a brief check by the physicist, in some cases, the need for this was eliminated by combining it with a routine visit (if the visit was due close to the move date), but in some instances, where the mobile was moved twice in quick succession, or the move fell at the midpoint between routine visits, this was not possible. Physics visits due to move checks alone over a 12 month period (January 91-December 91) amounted to 8 visits (average of one per mobile, per year), but there had been 24 moves in that period which had been combined with routine visits, ie. an average of 3 per mobile in total, although these were by no means evenly distributed.

The costs break down into four main areas, capital costs which are evaluated in table 5.5.2, daily QA costs which are in table 5.5.3, periodic testing shown in table 5.5.4 and administrative costs which are included in the summary table 5.5.1.

The following assumptions have been made in order to cost the two ways of working.

Number of processors: *10*
Number of X-ray units: *16 (8 mobile, 8 static).*
Number of centres: *7*

Assume a working pattern of *5 days per week*
46 weeks per year (accounts for Bank Holidays and moves)

Capital expenditure is defrayed over 10 years in equal amounts each year
Salary of a physicist or radiographer £20,000 per annum: 18p per minute

The new scheme cuts down on the physics checks to two per annum and makes the local radiographer responsible for checking the mobile X-ray units after a move. The daily checks performed by the radiographer are modified somewhat to provide more information.

The block check incorporates a lead block so that a sensitometric strip can be printed on one edge of the film. In addition, a test film of a stepwedge is taken to produce an X-ray sensitometric curve. The processor control sensitometry continues as before. Initially these results are sent to the physicist who will hold them centrally on computer. This may not be necessary later on once the new system has been proven but may still be useful for management reasons.
The proposed new system consists of two routine visits per year by the physicist for each unit, with the number and scope of the checks to be reduced in accordance with the observations made over the three year study. Radiographers would be responsible for move checks, but little additional time needs to be made available because at least one radiographer needs to be present when the unit is moved, so only the time spent doing the test and not the travelling time needs to be accounted for as an additional expense.

However, the physicist now requires approximately an additional hour per week, for 52 weeks, for data entry of the radiographers results unless they send them on a floppy. Eventually, when the system has been shown to be reliable, the radiographer's computer does all of the analysis. The physicist's role becomes that of a troubleshooting service plus what is effectively a calibration of the computer based system.

ITEM		OLD		NEW
	breakdowr	total	breakdowr	total
Capital p.a.		1340.00		3120.00
Annual checks (16 units)	240.88 X 16	3854.08		0.00
Routine checks (16 units)	233.82 X16 X2	7482.24	165.36 X 16 X 2	5291.52
Move checks (8 mobiles)	142.96 X 8	1143.68	30.81 X 24	739.44
Sensitometry (10 processors)	285.20 X10	2852.00	285.20 X 10	2852.00
Postage (7 centres)		0.00	11.50 X 52	668.25
Central record keeping		0.00	0.18 X 60 X 52	561.60
TOTAL		**£16,672.00**		**£13,232.81**

Table 5.5.1 Comparative costs of different QA regimes

It can be seen that there is a small financial saving to be made using the new system. This would be greater if the central record keeping was reduced or eliminated altogether. The main consideration has been that any changes must not put the image quality at risk.

CURRENT SYSTEM

Sensi-densi (x 8) @ £900:	£ 7,200
kV meter	£ 1,900
Dosemeter	£ 2,000
Alignment tool	£ 200
Slit	£ 700
Stars	£ 400
Jig	£ 200
Aluminium filters	£ 100
IQ test object	£ 700
	£13,400

PROPOSED SYSTEM

Dosemeter	£ 2,000
Alignment tool	£ 200
Slit	£ 700
Stars	£ 400
Jig	£ 200
Aluminium filters	£ 100
IQ test object	£ 700
Process control managers	£20,000
kV meters(7) @ £1900:	£13,300
Test objects (7) @ £700:	£ 4,900
	£44,500

or, excluding kV meters for each unit ***£31,200***

Table 5.5.2 Capital costs

FILM SENSITOMETRY

70p per film:	70p
3 minutes radiographer time:	54p
	£1.24 per processor, per day.
	= £285.20 per processor, per year.

OLD SYSTEM BLOCK CHECK

70p per film:	70p
13p per exposure:	13p
2 minutes radiographer time:	36p
	£1.19 per unit, per day.
	=£273.70 per unit, per year

NEW SYSTEM BLOCK CHECKS

70p per film:	70p
13p per exposure:	13p
4 minutes radiographer time:	72p
	£1.55 per unit, per day
	£356.50 per unit, per year

Table 5.5.3 Cost of Daily Checks

PHYSICIST

OLD SYSTEM		NEW SYSTEM	
ANNUAL CHECKS			
Time - 2 days:	£174.00	n/a	
Travel:	£ 25.00		
Films - 42 @ 70p:	£ 29.40		
X-ray - 96 @ 13p:	£ 12.48		
	£240.88		
ROUTINE CHECKS			
Time - 2 days:	£174.00	Time - 1½ days:	£130.50
Travel:	£ 25.00	Travel:	£ 25.00
Films - 36 @ 70p:	£ 25.20	Films - 10 @ 70p:	£ 7.00
X-ray - 74 @ 13p:	£ 9.62	X-ray - 22 @ 13p	£ 2.86
	£233.82		***£165.36***
MOVE CHECKS		(by Radiographer)	
Time - 1¼ days:	£108.75	Time - 2 hours	£ 21.60
Travel:	£ 25.00	Travel	£ 0.00
Films - 10 @ 70p:	£ 7.00	Films - 10 @ 70p	£ 7.00
X-ray - 17 @ 13p:	£ 2.21	X-ray - 17 @ 13p	£ 2.21
	£142.96		***£ 30.81***

Table 5.5.4 - Costs of Physics Checks

5.6 Monitoring daily measurements

The cumulative sum of variations from the mean (CUSUM), has been used in a range of areas of medicine for monitoring the performance of individuals against a specified target [148]. This technique uses one of the properties of random fluctuations; that they cancel each other out when averaged over a large enough sample (forty or so would be adequate). The size of the fluctuation is found by subtracting the measured value from the target value; in theory, the target value is identical with the mean, but in practice, there will be small errors associated with evaluating the mean and thus the target value may be slightly in error.

5.6.1 Interpretation of the CUSUM graph

If the CUSUM technique is used to monitor equipment which is in statistical control, the expected result is a graph which fluctuates around the zero value. Unlike the situation in which personal performance is being monitored, the equipment will not change its performance in response to the evaluation procedure, therefore the gradient should always be zero. This is only true if the correct target value was selected.

The first attempt to put CUSUM monitoring into practice resulted in a line of constant but non-zero gradient. The same result was seen for other units each with its own gradient. A brief consideration of the mathematics involved led to the conclusion that these results were due to small errors in the target value which accumulated in the CUSUM over time. The small errors should not have been a surprise, the target was set by taking the mean of the first ten measurements, too small a sample. When the targets were reset, the graphs started to show the expected behaviour.

Once the system was running properly, the graphs did wander away from zero from time to time and were shown to be very sensitive to small changes which would not have been so striking on a normal control chart. Small step changes resulted in a change of gradient and drifts resulted in a curve which appeared to be increasing exponentially. It was found that using the CUSUM in parallel with the run chart enabled accurate interpretations of the data to be made.

5.6.2 Computerised CUSUM

The next step, which has not yet been taken, is to use the computer to automatically calculate the local gradient on the CUSUM graph over a limited number of points, about four or five say. The following rules could then be used to act as a diagnostic indicator:

1. If the gradient is non-zero for two successive evaluations, then a change has occurred.

2. If the non-zero gradient maintains the same value, the change was a step change. The cause of the step change must be identified and remedied where appropriate, if the step change is not detrimental e.g. purchase of new cassettes, then the targets need to be reset.

3. If the non-zero gradient increases from one evaluation to the next, there has been an upward drift in the measured parameter.

4. If the non-zero gradient decreases from one evaluation to the next, there has been a downward drift in the measured parameter.

5. If the gradient becomes non-zero then returns to zero, a small step change occurred which was subsequently corrected; possible explanations are: a different cassette was used for a period of time, a different box of test film was used for a period, or the tests were done for a period of time at non-default settings.

It is relatively simple to incorporate a low level of artificial intelligence into the computerised measuring system which makes it possible to release the physicist from part of the role of trouble-shooter. Where further tests are indicated by the observed changes, then the physicist must still be called in. The reduction in the required physicists time will lead to an additional cost saving beyond that identified in table 5.5.1.

Chapter 6

Conclusions

6.1 Value of breast screening

Breast screening has been shown to be of some benefit in the reduction of mortality due to breast cancer [37,68,69,70,72,73], and indeed is one of the mainstays of the "Health of the Nation" initiative of the UK government. The costs associated with it are deemed unacceptably high by some analysts such as Skrabanek [19,30,149], who maintains that the money spent on breast screening could be much better spent on other aspects of health care. The fact that the cost per life saved is so high, is due substantially to the fact that breast screening in its current form is operating at the edge of the capabilities of technology. Genetic and biochemical markers hold the promise of a method of detecting breast cancer [58] even before it is evident using the most sensitive imaging techniques, but widespread use is several years away and the techniques are not yet proven.

The other limitation on the effectiveness of a screening programme in saving lives is the ability of the oncologist to treat cancer effectively. There are many forms of treatment available, surgical options include lumpectomy, simple mastectomy and radical mastectomy; chemotherapy is available either alone or in combination with surgery or radiotherapy and radiotherapy is available as an adjunct to surgery specifically to attempt to control metastatic disease. More recently, the drug tamoxifen has become widely used, it mimics hormones which are antagonistic to the growth of breast cancer. This drug has even been considered for prophylactic use in "high risk" women. There is a great need for large controlled trials of various permutations of the available techniques to establish which are the most effective for particular types of cancer.

6.2 Value of quality assurance

A mathematical model has been developed which attempts to describe how the number of false positive results and false negative results changes as a function of image quality. Simply moving the decision criteria decreases the number of false positive results at the expense of the false negative results. This can be a policy decision based on considerations of the consequences of either type of error. What image quality improvements are able to do is to change the shape of the distribution of results on which the radiologist is

working. This means that the number of false negative results *and* false positive results can be decreased simultaneously. The consequences of doing this are twofold, firstly the negative consequences associated with either type of error are reduced and secondly, there are considerable financial savings to be made by the reduction of numbers proceeding to the second and third stages of the screening process. The computer model which has been developed predicts from the initial screening statistics that a 10% improvement in image quality gives a 2% increase in the number of cancers detected, 8% decrease in false negatives and 32% reduction in false positives at screening. Results for this and for change of decision point are shown in table 6.2. This has been clearly illustrated by the Canadian study [82,83] which was widely criticised for the quality of its imaging and which concluded that breast screening was doing more harm than good in the 40 to 50 age group.

	Cancers detected	False negatives	False positives
Default	356	85	2733
IQ 10% decrease	349	93	3694
IQ 10% increase	363	78	1848
Decision point 10% further right	346	95	1998
Decision point 10% further left	363	79	3914

Table 6.2 The effect of changing image quality and decision point in the theoretical model

6.3 Optimising physical quality assurance

Quality assurance within the breast screening process is a very costly activity. If it can be streamlined in such a way that the final image quality and therefore the effectiveness of the programme, are not compromised, a financial benefit with no implications of reduction in detection rate can be achieved. Hand in hand with this objective is the aim to use the increased understanding of the ageing processes of the X-ray equipment to be able to predict tube failure and allow financial planning for replacement and scheduling of the replacement so as to cause the minimum disruption to the screening programme, for example, if it can be predicted that the tube has only six months of useful working life, a replacement can be timed to coincide with a bank holiday or some other period when screening would not ordinarily take place.

The data acquired has shown that measurements of output, whether made directly by the physicist or by using a proxy measure based on the automatic exposure device, are able to predict (with the benefit of hindsight) tube failure due to the tube becoming gassy. Catastrophic tube failure due to the filament going open circuit cannot be predicted by any measurements which can currently be made. Measurement of output alone does not guarantee good image quality, therefore a quality assurance programme measuring other parameters is still necessary.

It has been shown that changes of tube voltage within the normal working range (25 to 30 kV) produce no measurable effect on the image quality when assessed by test object suggesting that the $\pm$ 1 kV limit set in IPSM 59 [250] is unnecessarily strict, it will be amended in the forthcoming IPSM guidelines. This kind of kV control can be achieved easily by medium frequency generators, it is important to measure it, however, in order to make corrections to output measurements. Changes in optical density have a marked effect on the perceived image quality, the exact response to the nature of the change depends on the observer but is of the form that the image quality improves as the optical density increases, reaches an optimum value, which theoretically corresponds to the centre of the straight line portion of the characteristic curve, and then falls as the optical density increases further.

One screening unit within NE Thames had a processor on board the mobile. A fresh water supply was found to be necessary, the recirculation system gave problems if used for more than about six hours (depending on workload). The chemicals were removed from the tank of the processor prior to moving the mobile and it was found that it took 20 to 30 films to be processed for the processor to reach equilibrium, thereafter it performed no differently to any other processor. This equilibrium period corresponded to the time during which the physics checks were performed. It necessitated repeating some of the tests at the end of the day to ensure that the AEC function and the image quality had actually settled to a stable state.

6.3.1 Financial benefits of the new system of quality assurance

The cost of a single physics quality assurance visit can be approximated in the following way.

1. Consumables - films and chemicals

2. Wear and tear - use of X-ray equipment, processor and measuring equipment

3. Physicists time

4. Travel costs

The system proposed in chapter 5 reduces these costs but does not jeopardise image quality.

6.3.2 Quality benefits of quality assurance

The proposed system of quality assurance makes use of the go/no-go tests which the radiographers perform on a daily basis using blocks of perspex and the automatic exposure control of the X-ray machine to monitor the output of the X-ray machine. It has been demonstrated (in chapter 3) that if the AEC is functioning correctly, there is an inversely proportional relationship between the output and the mAs recorded from the block tests.

The statistical technique of the CUSUM makes these tests very sensitive to drifts in parameters, the CUSUM takes off in an exponential manner whereas for random noise, the fluctuations cancel out and the resulting function is a horizontal line. Care must be taken with this technique to distinguish between a drift and an error in the mean. The error in the mean shows up as a line of constant gradient whereas the drift shows up as either a positive or negative excursion away from the zero line. The outcome of the process is that an experienced physicist can evaluate the function of the X-ray tube on a weekly basis with none of the disruption to screening or the associated costs of a physics visit. The output of the tube is monitored much more frequently and corrective action can be taken more promptly implying that non-catastrophic faults will be dealt with more speedily. The equipment should work at its optimum level for a greater proportion of the time and quality of the images and consequently improvement in the cancer detection rate can be expected.

6.3.3 Organisational benefits of quality assurance

X-ray equipment is very heavily used in breast screening, and the ageing process is much accelerated, the lifetime of a tube is approximately three years compared to five or six years which would be expected from a standard diagnostic tube. Performance testing has been done on sixteen X-ray units at intervals of four months and has shown the ability, in retrospect, to predict the failure of a tube due to the presence of gas. Trends in some of the measured parameters, namely output and the post exposure mAs from the AEC block checks, have been identified which are apparent in X-ray tubes in the period leading up to failure which do not appear in other tubes which have continued to perform well.

This information has been used to produce a new regime of equipment testing which relies more heavily on the radiographers routine measurements and has reduced the amount of testing done by the physicist giving savings in terms of physicist time and less days when the equipment cannot be used for screening because it is undergoing testing.

There are some fault conditions which give no indication of their presence until a catastrophic failure occurs, in particular, the filament in the electron gun may go open circuit with absolutely no warning in the same way that the filament of a light bulb fails. No quality assurance measurement has been found which can predict this event.

There are other faults which can exist without causing the tube to fail but which cause the deterioration of image quality and therefore necessitate tube replacement. The alternative is to leave the tube alone until a drop in cancer detection rate indicates that there has been a loss in image quality. Such a procedure is very slow, it is difficult to tell whether the variations seen are normal fluctuations or real changes due to the drop in image quality even averaging over a period of one year. Long term deterioration has also been considered in the implementation of a new testing regime and has led to the conclusion that periodic physics testing cannot be eliminated entirely.

In addition, data produced by the radiographers on a daily basis has been incorporated into the analysis.

6.4 Further work

Some questions remain unanswered at the end of this project and suggest the need for further experimentation and development.

The observation of a decrease in output as an indicator of pending tube failure needs to be investigated more fully. This could be done

a) in a controlled trial using X-ray tubes which are not in routine use measuring the output as a function of time. This would allow an absolute value of output, based upon averages from many tubes, to be found which indicates that tube replacement is required.

b) by combining data on X-ray tube failure related to output measurements for all of the units within the national breast screening programme.

During the three years of experimentation one tube failed so early on that only three sets of measurements had been made, the tube at Southend was used for analysis and others were replaced for reasons like the focal spot going outside control limits, the results based on a single tube have limited applicability. The release of gas into the tube may well take place at different rates for different tubes, also, tubes from different manufacturers have rather different levels of output.

When a tube was replaced because of the focal spot size, no significant change in screening results was observed. It was assumed by Pritchard [3] that a resolution of 8.9 line pairs per mm was necessary for mammography, researchers in digital mammography [150,151] are of the opinion that 5 line pairs per mm is sufficient. Data from this piece of work supports that opinion, and standards need to be revised in the light of further controlled experiments to avoid unnecessary tube replacements.

The use of test objects to evaluate image quality has been shown to have limited value as an objective measure due to intra-observer variations, although at the present time they are the only practicable method for regular measurements. ROC analysis is the preferred method of objectively evaluating image quality but is too time consuming to be used within the screening programme.

During the course of this work some new test objects have become available;

i the TOR(MAM) from Leeds [152], which is a variation on the common theme of various shapes and sizes of object embedded within a tissue

equivalent matrix, it does however have an area which produces an image very similar to real breast tissue

ii a contrast detail test object from Holland [65] which requires the scorer to identify *where* the object is within a defined square as well as whether it can be seen, this aims to make the test more objective.

There is a need to develop test objects which are suitable for analysis by computer in order to reduce the degree of variability due to the scorer.

Differences in the positive predictive value for X-ray tubes with a tungsten anode compared to X-ray tubes with a molybdenum anode were observed in section 3.6.3, these were shown not to be statistically significant, but this may have been purely due to the limited numbers involved. Further investigation, perhaps using ROC analysis, would be able to separate out the effects of target material, processing and staff performance on the screening outcome.

REFERENCES

1 Department of Health and Social Security
Guidance notes for Health Authorities on mammographic equipment requirements for breast cancer screening STD/87/34
Department of Health, 14 Russell Square, London December 1987

2 Department of Health and Social Security 1986
Breast Cancer Screening. A report of a working group chaired by Professor Sir Patrick Forrest (HMSO, London)
ISBN 0-11-321071

3 J. Pritchard
Quality Assurance guidelines for mammography
National Health Service Breast Screening Programme, Institute of Physical Sciences in Medicine, Royal College of Radiologists, College of Radiographers (Joint publication)
NHSBSP Publications 1988

4 Institute of Physical Sciences in Medicine
The commissioning and routine testing of mammographic X-ray systems. Report No. 59,
Bocardo press, Didcot ISBN 0-904181-58-8

5 T. Roper, B. Adams.
Healthcare Data Briefing: Women and Cancer.
Health Service Journal, 23 Aug 1990, p1242.

6 A.Kalache.
Risk Factors for Breast cancer with Special reference to Developing Countries.
Health Policy and Planning 5:1,1-22.

7 J.M.G.Wilson & G.Jungner.
Principles and Practice of Screening for Disease.
World Health Organisation Public Health Paper 34. 1968.

8 S.W. Duffy.
Aggressiveness of Breast cancers Detected by Screening.
British Medical Journal 304:6838, 1377- 1378, May 23 1992.

9 R.D. Rubens.
Management of Early Breast Cancer.
British Medical Journal 304:6838, 1361-1364, May 23 1992.

10 A. O'Doherty.
Factors Achieving a High Pick Up Rate.
Symposium Mammographicum, 1992, p35.

11 B.A.Gusterson.
Prognostic Indicators.
Symposium Mammographicum 1992, p22.

12 J.Hayward.
An Overview of the DCIS problem.
Symposium Mammographicum 1992, p24.

13 L Tabar.
Breast cancer treatment and natural history - new insights from results of screening.
Lancet 339: 8790, 412-414. February 15th 1992.

14 S. Shapiro.
Periodic Breast cancer Screening in Seven Foreign Countries.
Cancer Supplement 69:7, 1919-1924.

15 B.M.Moores, et al.
Radiation Protection Associated with Well Women Breast Cancer Screening.
British Journal of Radiology 65(774): 552-553

16 L.Stanton, et al.
Dosage Evaluation in Mammography.
Radiology 1984(150):577-584

17 P. Cole, A.S.Morrison.
Basic Issues in Population Screening for Cancer.
Journal of the National Cancer Laboratory 64:5, 1263-1272, May 1980.

18 M.L.Brown.
Sensitivity Analysis in the Cost- Effectiveness of Breast Cancer Screening.
Cancer Supplement 69:7, 1963-1967.

19 P Skrabanek.
Breast Cancer Screening - a UK showdown.(Editorial)
British Journal of Breast Cancer Screening, Vol 40. December 1988.

20 M Harris.
The Doubts of a Dying Doctor.
Daily Mail Monday 11th June 1990.

21 TNM Classification of malignant tumours (4th Edition) 1987
International Union Against Cancer
ISBN 0-387-17366-8

22 T Nemoto, J Vana, RN Bedwani, HW Baker, FM McGregor, GP Murphy
Managment and Survival of Female Breast Cancer: Results of National Survey by the American College of Surgeons
Cancer 45:2917-24, 1980

23 IS Fentiman
Management options for ductal carcinoma in situ
Symposium Mammographicum Abstracts 1992, p24.

24 T.J.McMillan.
The Biology of Metastasis: An Overview.
Symposium Mammographicum 1992, p 24.

25 R.D.Rubens.
The management of Metastases.
Symposium Mammographicum 1992, p 25.

26 M.F.Spittle.
Radiotherapy in Breast Cancer.
Symposium Mammographicum 1992, p 25.

27 H.Joensuu, et al.
Histological features, DNA content and prognosis of breast carcinoma found incidentally or in screening.
British Journal of Cancer 64(3):588-592, September 1991

28 P.J. van der Maas.
The costs and effects of mass screening for breast cancer.
Rotterdam Department of Public Health and Social Medicine.
May 1988, ISBN 90-72245-03-2.

29 A.I.Mushlin, L Fintor.
Is Breast cancer screeening Cost Effective?
Cancer Supplement 69:7, 1957-1962.

30 P.Skrabanek.
The Case Against.
British Medical Journal 297, 971-972. 15 October 1988.

31 H Joensuu, et al.
Breast Cancer Found at Screening and Previous Detection by Women Themselves.
Lancet 339(8788):315. February 1st 1992

32 V.Champion.
The Role of Breast Self Examination in Breast Cancer Screening.
Cancer Supplement 69:7, 1985-1991. 1 April 1992

33 M.E.Constanza, et al.
Feasibility of Universal Screening Mammography - Lessons from a community intervention (Summary).
Archives of Internal Medicine 151(9):1851-1856, September 1991

34 R.S.Foster, J.K.Worden, M.C.Costanza, L.J.Solomon.
Clinical Breast Examination and Breast self-Examination.
Cancer Supplement 69:7, 1992-1997. 1 April 1992

35 D.P.Winchester.
Physical Examination of the Breast.
Cancer Supplement 69:7, 1947-1949. 1 April 1992

36 C.J.Baines.
Breast self-examination.
Cancer Supplement 69:7, 1942-1946. 1 April 1992

37 M.M.Roberts, F.E.Alexander, T.J.Anderson, U.Chetty, P.T.Donman, P.Forrest, W.Hepburn, A.Higgins, A.E.Kirkpatrick, J.Lamb, B.D.Muir, R.J.Prescot.
Edinburgh Trial of Screening for Breast Cancer : mortality at Seven Years.
Lancet 335 : 241 -246, 27 January 1990.

38 A.B.Miller, M.Tsechkovski.
Imaging Technologies in Breast Cancer Control.
Summary of a report of a World Health Organisation meeting.
American Journal of Roentgenology 148:1093-1094 June 1987.

39 G.U.V.Rao and P P Faburos.
Radiographic aspects of Xeroradiography.p 208-218. *in* The Radiologic Clinics of North America (editor Edward A Sickles) Vol.25 No. 5 Breast Imaging
WB Saunders, Philadelphia 1987 ISSN 0033-8389

40 P.B.Guyer.
An appraisal of Ultrasound in Breast Disease.
Symposium Mammographicum, 1992.

41 American Institute for Ultrasound in Medicine
Statement on mammalian in vivo ultrasonic biological effects.
December 1987

42 TA Whittingham
The safety of ultrasoud (review)
Imaging 6:33-51 1994

43 W. Dankiw.
Medical thermography. Australian Institute of Health: Health Care Technology Series No. 4 90/22318.

44 K.L.Williams, B.H.Phillips, P.A.Jones,S.A.Beaman, P.J.Fleming.
Thermography in screening for breast cancer.
Journal of Epidemiology and Community Health 44:2, 112-113, 1990.

45 A Moreno, et al.
Study of the ability of proton NMR spectroscopy of human plasma to differentiate between controls and breast cancer patients.
Oncology 50(2):110-5. March-April 1993.

46 H.G.Lewis-Jones, G.H.Whitehouse, S.J.Leinster.
The Role of MRI in the Assessment of Local Recurrent Breast Carcinoma.
Clinical Radiology 43: 197-204, 1991.

47 Hickman.
Differentiation of Scar Tissue and Carcinoma of the Breast using MRI with Gd DTPA Dynamic Enhancement.
Symposium Mammographicum 1992, p 31.

48 C.Westbrook.
Magnetic Resonance Imaging of the Breast.
Radiography Today 58:663, 15-17 August 1992.

49 S.H.Heywang-Kobrunner
Diagnosis of breast cancer with magnetic resonance. Review after 1250 patient examinations.
Electromedica 61(2):43-52 , 1993

50 A.Del Maschio and A.Vanzulli.
Digital Chest Images : Evaluation of Normal Structures.
Medical Review 35: 12-16.

51 U.Bick, W.Wiesman, H.Lanzen, M.Fiebich, H.J.Vonhengeke, P.E.Peters.
Utilising digital luminescence Radiography in Paediatric Radiology: a Report of Initial Experiences.
Electromedica 59(1):26, 1991.

52 A R Cowen, D S Brettle et al.
A Preliminary investigation of the imaging performance of photo stimulable phosphor computed radiography using a new design of mammographic QC test object.
British Journal of Radiology 65(774):528-535. June 1992.

53 GJS Parkin.
Direct Digital Mammography.
Symposium Mammographicum 1992, p 28.

54 GJS Parkin
Digital mammography- The technique of the future
RAD 20(230):18 July 1994

55 D.H.Davies and D.R.Dance.
Computer Interpretation of Digital Mammograms.
Symposium Mammographicum 1992, p 41.

56 EJ Silva
Clinical case studies of breast cancer studies with CT
Electromedica 59(4): 106-109, 1991

57 P A Garret, et al.
HRAS Protooncogene Polymorphism and Breast Cancer (ABS).
Cancer Epidemiology, Biomarkers and Prevention 2(2): 131-8.
March-April 1993.

58 T J Powles.
Screening of inherited breast cancer with DNA markers.
Lancet 341:8857, 1422. May 29th 1993.

59 P Lalle et al.
Screening of inherited breast cancer with DNA markers.
Lancet 341: 8857, 1422. May 29 1993.

60 H W Simpson.
A breast pre-cancer test? Preliminary results based on a breast temperature rhythym abnormality during the menstrual cycle.
Breast Cancer Research and Treatment 16:51-55. 1990.

61 HW Simpson and K Griffiths
The diagnosis of breast pre-cancer by the chronobra I Background review II The breast cancer pre-test
Chronobiology International 6(4): 355-393 1989

62 R.McLelland, et al.
The Amercan College of *Radiology* Mammography Accreditation Program.
American Journal of Roentgenology 157(3):473-479

63 R.M.Adamson.
Screening in the West Down Under.
Symposium Mammographicum 1992, p 40.

64 J G Murray et al.
Assessment of mammographic film processor performance in a hospital and mobile screening unit
British Journal of Radiology 65(780):1097-1101

65 M Thijssen
Personal Communication December 1992

66 Martin Vessey and Muir Gray.
Breast Cancer Screening 1991: Evidence and Experience since the Forrest Report.
NHSBSP Publications. ISBN 1 871997 06 2. January 1991.

67 S.Shapiro, W.Venet, P.Strax. L.Venet, R.Roesner.
Ten to Fourteen Year effect of Screening on Breast Cancer Mortality.
Journal of National Cancer Institute 69(2): 349-355 August 1982.

68 S. Shapiro.
Evidence on Screening for Breast cancer from a Randomised Trial.
Cancer 39:2772-2782, 1977.

69 L.Tabar, A.Gad, L.H.Holmberg, U.Ljungquist, C.J.G.Fagerberg, L.Badetorp, O.Grontoft, B.Lundstrom, J.C.Maison, G.Eklund, N.E.Oly, F.Petterson.
Reduction in Mortality from Breast Cancer after Mass Screening with Mammography.
Lancet 1: 829-83. 13 April, 1985.

70 I.Andersson, K.Aspegren, L.Janzon, T.Landberg, K.Lindholm, F.Linell, O.Ljingberg, J.Ranstam, B.Sigfusson.
Mammographic screening and mortality from breast cancer: the Malmo mammographic screening trial.
British Medical Journal 297: 943-948, 15 October 1988.

71 L Nystrom.
Breast cancer screening with mammography: overview of Swedish randomised trails.
Lancet 341(8851): 973-8. April 17 1993.

72 A.L.M.Verbeek, J.H.C.L.Hendricks, et al.
Reduction of Breast Cancer Mortality through Mass Screening with Modern Mammography (Nijmengen).
Lancet : 1222-1224. June 2nd 1984.

73 A.L.M. Verbeek, J.H.C.L. Hendricks, R. Holland, F. Sturmans.
Mammographic Screening and Breast cancer Mortality: Age Specific Effects in Nijmegen Project 1975-1982 (letter).
Lancet:865-866, April 13 1985.

74 UK Trial of Early Detection of Cancer Group.
Trial of early detection of Breast Cancer: description of method.
British Journal of Cancer 44, 618-627, 1981.

75 L.G.Kessler.
The Relationship Between Age and Incidence of Breast Cancer.
Cancer Supplement 69:7, 1896-1903.

76 C.Mettlin.
Breast Cancer Risk factors.
Cancer Supplement 69:7, 1904-1910.

77 B.A.J.Ponder (letter).
Breast Cancer Genes.
British Medical Journal 302: 861-862 13 April 1991.

78 C M Frienden Reich, et al.
A Cohort Study of Alcohol Consumption and Risk of Breast Cancer (ABS).
American Journal of Epidemiology 137(5): 512-20. 1st March 1993.

79 M.P.Vessey.
HRT and Breast Cancer.
Symposium Mammographicum 1992, p 21.

80 S.A.Feig.
A New Method for Assessment of Radiation Risk from Screening Mammography.
Recent Results in Cancer Research 119:141-150, 1990

81 M.E.Costanza.
Breast Cancer Screening in Older Women.
Cancer Supplement 69:7, 1925-1931.

82 D.B.Kopans.
Breast Screening in Women Under 50.
Lancet 338:8764, p447, August 17th 1991.

83 P. Last.
Breast screening for the Under 50's.
Lancet 338:876, p583 August 31st 1991.

84 K Heilbrunn
Younger Women benefit from mammography
Report from RSNA, Chicago 1994
RAD magazine March 1994

85 J Kitawaki, T Kim, H Kanno, et al.
Growth Suppression of MCF-7 cancer cells by aromatase inhibitor screening.
Journal of Steroid Biochemistry and Molecular Biology 44(4-6): 667-70. March 1993.

86 D V Schapira, N B Kumar, G H Lyman.
Variation in body fat distribution and breast cancer risk in the families of patients with breast cancer and control families.
Cancer 71(9): 2764-8. May 1st 1993.

87 EF Conant, RL Dillon, J Palazzo, SM Ehrlich, SA Feig
Imaging findings in mucin containing carcinomas of the breast: correlation with pathologic features
American Journal of Roentgenology 163:821-824 1994

88 I.C.Henderson.
Biologic Variation of Tumours.
Cancer Supplement 69:7, 1888-1895.

89 M.J.Garton, et al.
Recruitment Methods for Screening Programs - trial of a new method within a regional osteoporosis study (Summary).
British Medical Journal 305(6845):82-84, 11 July 1992

90 T Hoare
Response of older women to invitations for breast screening (letter)
British Medical Journal 302:181, 19 January 1991.

91 J S Spratt, J Zivnheld, J M Yancey.
Breast cancer detection demonstration project data can determine whether the prognosis of breast cancer is affected by the time of surgery during the menstrual cycle.
Journal of Surgical Oncology 53(1): 4-9. May 1993.

92 RA Badwe, WM Gregory, MA Chauday, MA Richards et al
Timing of Surgery during menstral cycle and survial of premenopausal women with operable breast cancer.
Lancet 337:1261-4, 1991

93 SM Sagar, P Lopez
Treating elderly patients with breast cancer (letter)
British Medical Journal 304:1376-1377 23 May 1992

94 K Dookeran, A Stotter, R Windle, R Walker
Treating elderly patients with breast cancer (letter)
British Medical Journal 304:1376-1377 23 May 1992

95 Z Rayter, R F Phipps.
Primary Medical Treatment in Breast Cancer.
British Medical Journal, 302. 5th January 1991.

96 R Amalvic, F Santamaria, F Roberts, J Seigle, C Allshculer, JM Kuntz, JM Spitalier, H Brandore, Y Ayme, JF Pollet, R Burmeister, R Abed
Radiation therapy with or without primary limited surgery for operable breast cancer. A 20 year experience at the Marseille Cancer Institute
Cancer 49:30-34, 1982

97 RW Blamey (Chair)
Quality assurance and the assurance of quality in the management of caners detected in breast cancer, the role of BASO
European Journal of Surgical Oncology
16:532-535, 1990

98 D Ingram R Block (editors)
Chapter 3 Imaging
Mathematical methods in medicine: Part 2 Applications in clinical specialties
Wiley 1986
ISBN 0 471 90046 x

99 J A A M van Dijck.
One view versus two view mammography in baseline screening for breast cancer: a review.
British Journal of Radiology 65(779): 971-976, 1992

100 K Young personal communication January 1993

101 D.K.Owens, H.C.Sox.
Chapter 3 Medical Informatics
Medical Decision Making: Probabilistic Medical Reasoning, editors EH Shortliffe and LA Perrault
Addison Wesley 1990
ISBN 0 201 06741 2

102 HE Johns JR Cunningham
The physics of radiology, 4th edition
Thomas Books, Springfield, Illinois, 1983
ISBN 0 398 04669 7

103 PH Carter (editor)
Chesneys' equipment for student radiographers. Chapter 11.
Blackwell Scientific Publications 1994
ISBN 0 632 02724 X

104 S.A.Beaman, S.C.Lillicrap.
Optimum X-ray spectra for Mammography.
Physics in Medicine and Biology 27(10):1209-1220 October 1982.

105 S.A.Beaman, S.C.Lillicrap, J.L.Price.
Tungsten Anode Tubes with K-edge Filters for Mammography.
British Journal of Radiology 56:721-727 1983.

106 R.J.Jennings, R.J.Eastgate, M.P.Siedband, D.L.Ergun.
Optimal X-ray spectra for Screen Film Mammography.
Medical Physics 8(5):629-639, September/October 1981.

107 D Dance
Personal Communication 1991

108 R.A.Wilson.
Rare Earth Screens Versus Calcium Tungstate Screens.
Radiology 147:288-9, April 1983.

109 T.Fearon, J.Vucich, J.Hoe, W.J.McSweeney, B.M.Potter.
A Comparative Evaluation of Rare Earth Screen Film Systems, System Speed, Contrast, Sensitometry, RMS Noise, Square, Wave Response Function and Contrast - Dose Detail Analysis.
Investigative Radiology 21:654-63, August 1986.

110 W.K.Chu, W.Sangster, C.A.Dobry.
Comparison of radiation exposures between Conventional and Rare-earth Screen Film Systems.
Health Physics 49(5):958-961, November 1985.

111 WR Hendee, R Ritenour
Medical Imaging Physics (3rd edition)
Mosby Year Book, St Louis, Mo 1992
ISBN 0-8151-4241-2

112 DJ Dubowitz, PD Britton
Image guided breast biopsy
RAD 20(230):13-14 July 1994

113 User Manual of X-ray Quality Control Procedures in Mammography.
North East Thames Regional Health Authority, 1990.

114 D Schooling
Personal Communication 1993

115 J Law
UK Mammography Physics Group news letter
Private Communication, February 1990

116 J Law.
The influence of focal spot size on image resolution and test phantom scores in mammography.
British Journal of Radiology 66 (785), p 441-446. May 1993.

117 J Law.
Measurement of focal spot size in mammography X-Ray tubes.
British Journal of Radiology 66 (781), p 44-50. 1993

118 British Standards Institution
British Standard methods for determining the characteristics of focal spots in diagnostic X-ray tube assemblies for medical use BS 6530
BSI, London 1984

119 G.R.Hammerstein, et al.
Absorbed Radiation dose in Mammography.
Radiology 130:485-491 February 1979.

120 D.Dance.
Monte Carlo calculation of conversion factors for the estimation of mean glandular breast dose.
Physics in Medicine and Biology 35:9, 1211-1219, 1990

121 Objectives for the breast screening programme
NHSBSP internal publication
January 1993

122 National Radiological Protection Board, Health and Safety Executive and Health Departments 1988
Guidance notes for the protection of persons using ionising radiations arising from medical and dental use
HMSO, London 1988
ISBN 0 85951 299 1

123 NHS Procurement Directorate
Technical requirements for the supply and installation of equipment for diagnostic imaging and radiotherapy (1989)
TRS 89

124 WE Deming
Out of the crisis:quality, productivity and competitive position
Cambridge University Press 1986
ISBN 0 91137901 0

125 Protection Against Ionising Radiation from External Sources Used in Medicine.
International Comission on Radiological Protection, Publication 33, 1982.

126 The Ionising Radiation (Protection Of Persons Undergoing Medical Examination or Treatment) Regulations 1988. Statutory instrument No.778
HMSO, London 1988
ISBN 0 11 0867785

127 The Ionising Radiation Regulations 1985. Statutory instrument No. 1333
HMSO, London 1985
ISBN 0 11 057333 1

128 Approved code of practice: the protection of persons against ionising radiation arising from any work activity 1985
HMSO, London
ISBN 0 11 883838 5

129 Royal College of Radiologists and NRPB.
Patient Dose Reduction in Diagnostic Radiology.
Documents of the NRPB Vol.1.
London HMSO 1990.
ISBN 0 85951 327 0

130 Quality Criteria for Good Radiographic Images.
Working Document 2nd Edition, June 1990.
CEC Study Group.

131 C.Hessler, Christian Depeursinge, Mikail Grecescu, Yvan Pochon, Sergio Raimondi and Jean Franscois Valley.
Objective Assessment of Mammography Systems Part I: Method .
Radiology 156(1): 215-219, July 1985.

132 C.Hessler, Christian Depeursinge, Mikail Grecescu, Yvan Pochon, Sergio Raimondi and Jean Franscois Valley.
Objective Assessment of Mammography Systems Part II:Implementation.
Radiology 156(1): 221-225, July 1985.

133 J.C.P. Heggie et al.
ACPSEM Position Paper: A Quality Assurance Programme for Mass Screening in Mammography.
Australian Physical and Engineering Scinces in Medicine 12:4, 252-259, 1989

134 K.T.K.Karila.
Quality Control of Mammographic Equipment: a Five year Follow-up.
British Journal of Radiology 61: 1155-1167, 1988.

135 A Burch
Private communication , June 1990

136 A LI Evans
The evaluation of medical images
Medical Physics Handbook 10
Adam Hilger 1981
ISBN 0 85274 518 4

137 C.P.McDonagh, J.K.Leake, S.A.Beaman.
Optimum X-ray spectra for Mammography : Choice of K-edge Filters for Tungsten Anode Tubes.
Physics in Medicine and Biology 29(3): 249-252, 1984

138 M Whall.
Optimisation of X-Ray spectra for mammography.
British Journal of Radiology 66(4):384-385. April 1993.

139 H. Aichinger, S. Joite-Barfus and P. Marhoff.
Automatic Exposure Control system in Mammography.
Electromedica 58(2):61-66, 1990

140 C Kimme-Smith
"Mammography screen-film selection, film exposure, and processing" from Screen Film Mammography, Imaging Consideration and Medical Physics Responsibilities:135-158
Proceedings of SEAAPM Spring Symposium April 6, 1990

141 R.W.Gurney, N.F.Mott.
The Theory of Photolysis of Silver Bromide and the Photographic Latent Image.
Proceedings of the Royal Society (London) 164:151-167 January 1938.

142 JW Mitchell
The formation of latent image in photographic emulsion grains (article)
Photographic Science and Engineering 25(5):170-188 1981

143 Presentation B King
Fuji Training Day February 1993

144 B Hall
Private Communication 1992

145 Fuji training course notes - personal communication
16th February 1993

146 K Ishikawa
What is total quality control? The Japanese Way
Prentice Hall
Englewood Cliffs NJ, 1985
ISBN 0 13952433 9

147 Kodak Patent.

148 S.M.Williams, B.R.Parry, M.M.T.Schlup.
Quality Control:an application of the CUSUM.
British Medical Journal 304:6838, 1359-1361, 23 May 1992.

149 P Skrabaneck.
Breast Cancer Screening - the current position.
British Medical Journal 302: 6789, 1401. June 8th 1991.

150 DS Brettle et al.
Computed radiography in mammography
RAD 20(230):15 July 1994

151 N Karssemeijer et al.
Spatial resolution in digital mammography
Investigative Radiology 28(5): 413-419 May 1993

152 AR Cowen DS Brettle NJ Coleman and GJS Parkin
A preliminary investigation of the imaging performance of photo stimulable phosphor computed radiography using a new design of mammographic quality control test object
British Journal of Radiology 65:528-535 1992

BIBLIOGRAPHY

Cervical/uterine cancer

G Johannesson, G Geirsson, N Day.
The Effect of Mass Screening in Iceland 1965-74 on the incidence and mortality of cervical cancer.
International Journal of Cancer 21:418-425 (1978).

A B Miller, T Visentia, E R Howe.
The effect of hysterectomies and screening on mortality from cancer of uterus in Canada.
International Journal of Cancer 27:651-657 (1981).

N Takenaga, I Kai, G Ohi.
Evaluation of three cervical cancer detection programs in Japan with special reference to cost benefit analysis.
Cancer 55: 2514-2519, 1985.

J B Thorn, et al.
Cost of detecting and treating cancer of the uterine cervix in North East Scotland in 1971.
Lancet, p 674-676. March 22, 1975.

Database

J A Hokanson, et al.
Design considerations for a Medical School Hospital Cancer Patient Data System.
Cancer 51: 1556-1561, 1983.

Decision Making

KS Bay, D Flathman
The worth of a screening programme:An application of a statistical decision model for the benefit evaluation of screening projects
American Journal of Public Health 66:145-150 February 1976

DG Brown, RF Wagner and MS Pastel
Applications of information theory to the assessment of the performance of radiological imaging systems p156-161
in The Radiologic Clinics of North America (editor Edward A Sickles)
Vol.25 No. 5 Breast Imaging
WB Saunders, Philadelphia 1987
ISSN 0033-8389

KA Eagle
Medical decision making in patients with chest pain (letter)
New England Journal of Medicine 324(18):12872-1283 May 2 1991

T H Lee, et al.
Ruling out acute myocardial infarction
*New England Journal of Medicine,*324(18):1239-1246. May 2nd 1991.

HC Sox
Probability Theory in the use of diagnostic tests
Annals of Internal Medicine 104:60-66 1986

Diagnosis

A.Yelland et al.
Diagnosing Breast Carcinoma in Young Women.
British Medical Journal 302: 618-620, 16 March 1991.

Digital Imaging

Medical Devices Directorate
An evaluation of the Siemens (DLR) Digital Luminescence radiography systems
MDD/91/28, Department of Health 1991
ISBN 1-85197-685-X

Medical Devices Directorate
A clinical evaluation of the Philips Graphics II computed radiography system
MDD/91/09
Department of Health 1991
ISBN 1-85197-596-9

R D Muller, et al.
Interactive image post-processing of digital luminescence radiographs.
Electromedica 58(2):48-55, 1990.

DNA testing

T. Mori, H. Inaji, H Koyama, R Abe, M Nihei, M Izuo, T Ogawa, K Enomoto, H Sato, F Kasumi et al.
Evaluation of an improved dot-immunobinding assay for carcinoembryonic antigen in nipple discharge in early breast cancer (abstract)
Japanese Journal of Clinical Oncology 22(6): 371-376, December 1992

Dose considerations

G.T.Barnes, D.P.Chakroborty.
Radiographic Mottle and Patient Exposure in Mammography.
Radiology 145:815-821, December 1982.

D.J.Brenner, H.I.Amols.
Enhanced Risk from Low Energy Screen/film Mammography X-rays.
British Journal of Radiology 62:910-914, 1989.

G.Cohen, L.K.Wagner, D.L.McDaniel, L.H.Robinson.
Dose Efficiency of Screen Film Systems used in Paediatric Radiography.
Radiology 152:187-193 July 1984.

D Dance and G Day
Proceedings of seminar on patient exposure in medical X-ray diagnosis, Munich, Germany, April 1981
EUR 7438 : 227 -244

A Ferriman.
Through a plate darkly.
The Times. 25th May 1993.

M.Fitzgerald, D.R.White, E.White, S.Young.
Mammographic Practice and Dosimetry in Britain.
British Journal of Radiology 54:212-220, March 1981.

D.G.M.Greig.
The Reduction of X-ray Exposure in Dental Practice using Rare Earth Screen/Fast Film Combinations.
British Dental Journal 155:17-18, 9 July 1983.

D Hart, J C Le Heron.
The Distribution of Medical X-Ray Doses amongst individuals in the British population.
British Journal of Radiology 65 No 779, 996-1002, 1992

J Law.
Calibration of ion chambers for use in mammography.
British Journal of Radiology 66 (781), p 51-54. 1993

J.Law.
Patient Dose and Risk in Mammography.
British Journal of Radiology 64:360-365, 1991.

D S MacDonald-Janowski, C P Lawinski.
The effect of thin K-edge filters on radiation dose in dental radiography.
British Journal of Radiology 65 (779)M O 990-995. November 1992.

C J Martin, et al.
Implementation of a programme for reduction of radiographic doses and results achieved through increases in tube potential.
British Journal of Radiology 66(3): 228-233. March 1993.

B M Moores, E T Henshaw.
Radiation protection associated with well women breast cancer screening
British Journal of Radiology 65:774. June 1992.

P Morgan, M Chevalier, E Vaño
Comparative study of dose values and image quality in mammography in the area of Madrid
British Journal of Radiology 67:556-563, 1994

H.D.Nagel
NIFE - (Filtration measurement)
Physics in Medicine and Biology 35(a): 1335-1343, September 1990.

NCRP Report No.85, March 1986.

J Patnick.
Radiation protection associated with well woman breast cancer screening.
British Journal of Radiology 65:778, 949-950. October 1992.

K J Robson, C J Kotre, K Faulkner.
Latent image fading and mean glandular dose in mobile mammography screening.
British Journal of Radiology 65(780):1148-1149. December 1992.

M Rosenstein.
Mathematical model for breast doses in mammography.
in The Radiologic Clinics of North America (editor Edward A Sickles)
Vol.25 No. 5 Breast Imaging
WB Saunders, Philadelphia 1987
ISSN 0033-8389

AM Tucker
Breast Screening Risks
Times 21 May 1990
A. Walker.
Patients and X-rays.
British Medical Journal 302: 71-72, 12 January 1991.

M Whall, EA McNeil
Radation dose to the breast from screening mammography in the West Midlands
British Journal of Radiology (Supplement) 1992

Equipment

AR Cowen, DS Brettle, A Workman
Technical note: Compensation for field non-uniformity on a mammographic X-ray unit
British Journal of Radiology 66:150-154, 1993

D Cripps.
Use of Macro-radiography in the initial examination of the Petrous Temporal Region.
Radiography Today 57(650). July 1991.

Department of Health, Medical Devices Directorate
Revised guidance notes for Health Authorities on mammographic X-ray equipment requirements for breast cancer screening
STD/90/46 November 1990
ISBN 1 85197 582 9

K .Faulkner, K.J. Robinson, C.J.Kotre.
X-ray Mammography and Breast Compression (letter).
Lancet 340, 797-789, Sept 26, 1992.

Health and Safety Executive
Fitness of equipment used for medical exposure to ionising radiation
Guidance note PM77 May 1992
HMSO, London ISBN 0 11 886313 4

CJ Kotre, KJ Robson, K Faulkner
Technical note: Assessment of X-ray field alignment in mammegraphy
British Journal of Radiology 66:(782): 155-175, 1993

M.V.Merrick.
Evaluation of Skeletal Metastases.
British Journal of Radiology 65:777, 803-806, 1992.

M. Sangheera, W. Peh.
Breast ductography in the investigation of Nipple Discharge.
Radiography Today 57:650, 15-16, 1991.

A Schentz
Automatic Exposure Control for Mammography
Electromedica 60(3): 91-92, 1992

D.J.Watmough, K.M.Quan.
X-ray Mammography and Breast Compression.
Lancet 340:122, 11 July 1992

Films and intensifying screens

B.A. Arnold, H.Eisenberg, B.E. Bjarngard.
Measurement of Reciprocity Law Failure in Green Sensitive X-ray Films.
Radiology 126:493-498, February 1978.

B.A.Arnold, E.W.Webster, L.Kalisher.
Evaluation of Mammographic Screen-film Systems.
Radiology 129: 179-185, 1978.

Y.Y.Bakir, E.J.Roebuck, D.Whelpton, B.S.Worthington.
Film Screen Combinations for Mammography (letter).
British Journal of Radiology 57:653 July 1984.

K Doi , G. Holje, L.N.Loo and H.P.Chan.
Evaluation of Resolution Properties of Radiographic Screen film systems.
in The Radiologic Clinics of North America (editor Edward A Sickles)
Vol.25 No. 5 Breast Imaging
WB Saunders, Philadelphia 1987
ISSN 0033-8389

Derek Gifford.
A Handbook of Physics for Radiologists and Radiographers.
Table 42.1,
Wiley & Sons 1984.
ISBN 0 471 90172 5

A.G.Haus.
Trends in Screen Film Mammography.
Eastman Kodak Company, 1986.

E.T.Henshaw.
Measurements of Response of some Film Screen Combinations.
Internal KCH, June 1975.

Y.Higashida, P.H.Frank, K.Doi.
High Speed, Single Screen/Single Emulsion Film, Systems:
Basic Imaging Properties and Preliminary Clinical Applications.
Radiology 149:571-577, November 1983.

A.E.Kirkpatrick, J.Law.
A Comparative Study of Films and Screens for Mammography.
British Journal of Radiology 60:73-78, January 1987.

Y.Kodera, K.Doi, H.P.Chan.
Absolute Speeds of Screen Film Systems and their Absorbed Energy Constants.
Radiology 151:229-236, April 1984.

A.S.Murray, B.Heaton.
Comparison of various Mammographic Film Screen Combinations with regard to Dose Absorbed by the Breast.
Proceedings of 3rd International Symposium on Radiological Protection. Advances in Theory and Practice. 169-174, 1982

D.Paix, A.F.Holloway.
Sensitivity, Resolution and Contrast of Film Screen Combinations.
Australasian Radiology 19:71-73, February 1985.

L.J.Price, J.Gamble, P.Pearce.
Film Screen Combinations in Mammography (letter).
British Journal of Radiology 58:99-100, January 1985.

E.J.Riihimäki, O.A.Korhola, H.T.Suoranta, M.K.Valle.
Improvement of X-ray Intensifying Screen Efficiency by Special Design of the Screen.
Radiology 142(1): 229-31, January 1987.

W.L.Smith, E.A.Franken, S.M.Frangi, C.Windsor.
Selection of Optional Screen Film Combination for use in the Neonatal Intensive Care Unit.
American Journal of Roentgenology 139:1051-1053, December 1982.

C.J.Vyborny, C.E.Metz, K.Doi, K.Rossmann.
Screen/film System Speed: its Dependance on X-ray Energy.
Radiology 125: 811-816, 1977.

Guidelines

CEC study group
European Guidelines for QA in Mammography Screening.
Draft 1990

NHS Breast Screening Programme
Quality Assurance Guidelines for Surgeons in Breast Cancer Screening
NHSBSP Publications No. 20, May 1992

NHS Breast Screening Programme
Information and advice for radiographers
NHSBSP publications April 1993
ISBN 1-871997-61-5

R Warren.
Breast screening applies for BS5750/ISO 9002.
BIR Bulletin. September 1992.

Image analysis

CJ Kotre, KJ Robson, K Faulkner
Measurements of the frequency distribution of optical density in screening mammography
British Journal of Radiology 67(9):856-859 September 1994

P Taylor, S Hajnal, M-H Dilhuydy, B Barreau
Measuring image texture to separate "difficult" from "easy" mammograms
British Journal of Radiology 67(5):456-463 May 1994

Image Quality

B.A.Arnold, H.Eisenberg, B.E.Bjarngard.
The LSF and MTF of Rare-earth Oxysulphide Intensifying Screens.
Radiology 121: 473-477, 1976.

J.A.Bencomo, B.G.Fallone.
A Logit Model for the MTF of Screen Film Systems.
Medical Physics 13(6):857-860, November/December 1986.

P.C.Bunch.
Comparison of S/N Ratios for some new Screen/film Systems.
SPIE :Medical Imaging and Instrumentology 486 : 99-107, 1984.

H.P.Chan, K.Doi.
Studies of X-ray Energy Absorption and Quantum Noise Properties of X-Ray Screens by use of Monte Carlo Simulation.
Medical Physics 11(1):37-46, January/February 1984.

G.Cohen, D.L.McDoniel, L.K.Wagner.
Analysis of Contrast - Detail experiments.
Medical Physics 11(4):469-473, July/August 1984.

L.Desponds, et al.
Image Quality Index for Screen/film Mammography.
Physics in Medicine and Biology 36(1): 19-33, 1991.

C.E.Metz, C.J.Vyborny.
Werner Spectral Effects of Spatial Correlation between the sites of Characteristic X-ray Emission and Reabsorption in Radiographic Screen Film Systems.
Physics in Medicine and Biology 28(5):547-564 May 1983.

R.M.Nishikawa, K.J.Yaffe.
Signal to Noise Properties of Mammographic Film Screen Systems.
Medical Physics 12(1):32-39, January/February 1985.

J M Sandrik, R J Jennings, R F Wagner.
Comparison of MTF, noise power spectrum and sensitometric measurements of X-ray screen-film systems made in two different laboratories.
p162-165 *in* The Radiologic Clinics of North America (editor Edward A Sickles)
Vol.25 No. 5 Breast Imaging
WB Saunders, Philadelphia 1987
ISSN 0033-8389

Siemens Test Object (SIB) Handbook

D R White and R J Martin
Epoxy resin based tissue substitutes
British Journal of Radiology 50:814-821, 1977

D.R.White, A.K.Tucker.
A test Object for Assessing Image Quality in Mammography.
British Journal of Radiology 53:331-335, 1980.

MRI

RW Kerslake
Magnetic resonance mammography (review)
RAD magazine 19(214):15 March 1993

RC Smith
MR Mammography
MR Clinical Symposium 6(1) 1992

Photography

D.N.Chesney, M.O.Chesney.
Radiographic Imaging
4th Edition.
Blackwell Scientific Publications 1981.
ISBN 0-632-00562-9

D.Jenkins.
Radiographic Photography and Imaging Processes.
MTP Press Limited 1980.
ISBN 0 85200 208 4
J Law, A Kirkpatrick
Film processing for mammography
British Journal of Radiology 61(730):939-942, October 1988

D Mc Lean
An evaluation of the effect of processing conditons on mammographic film contrast, fog levels and speed
Australasian Radiology 36(3): 234-237, August 1992

DP Roberts, NL Smith
Radiographic Imaging- A pactical approach
Churchill livingstone 1988
ISBN 0-443-03061-8

A. E. Saunders.
Reciprocity Failure.
Journal of Photographic Science 39:85-94, 1991.

Prognosis

I R Brand, D A Sapherson and T S Brown.
Breast Imaging in women under 35 with symptomatic breast disease.
British Journal of Radiology 66(785):349-397. May 1993.

RM Clark
Prognosis of breast cancer associated with pregnancy (letter)
British Medical Journal 302:1401 8 June 1991

J M Dixon and T G John.
Morbidity after breast biopsy for benign breast disease in a screened population.
Lancet 339(8785):128, 1992.

P J Klemi, H Joensuu, S Toikkanen.
Aggressiveness of Breast Cancers Found with and without Screening.
British Medical Journal 304(6825): 467-469. Feb 22 1992

R F Mould.
Cancer Statistics.
Adam Hilger, Bristol 1983. ISBN 0-85274-541-9.

Scatter

M Friedrich
The usefulness of a moving grid in mammography (letter)
British Journal of Radiology 59:1145-1146, November 1985

SF Keevil, CP Lawinski, EJ Morton
Measurement of the physical charcteristics of anti-scatter grids (private communication)

A.E.Kirkpatrick, J.Law.
The Usefulness of a Moving Grid in Mammography.
British Journal of Radiology 58:257-258 1985 March.

Screening

V Beral.
Breast Cancer -Mammographic Screening.
Lancet 341(8859): 1509-1510. 12 June 1993.

L. Breslow.
Review and Future perspectives of Cancer Screening Programs
Proceedings of Third International Symposium on the Detection and Prevention of Cancer.
New York. 26 Apr-1 May 1976, p1177-1209.

J. Chamberlain.
Specificity of Screening in the United Kingdom trial of early detection of breast cancer.
British Medical Journal 304, p346-349.

JM Elwood, B Cox and A K Richardson.
Breast Cancer Screening with Mammography (letter).
Lancet 341(8859):1531, 1993.

M Hakama.
Breast Cancer Screening with Mammography (letter).
Lancet 341(8859):1531, 1993.

H.J.de Koning, et al.
Advanced Breast Cancer and its Prevention by screening.
British Journal of Cancer 65(6):950-955 1992

D B Kopans.
Mammography (letter).
Lancet 341 (8850):957. April 10 1993.

Lancet noticeboard.
Canadian Study of Breast Screening Under 50.
Lancet 339: 8807, 1473-1474. June 13 1992.

T Morimoto, et al.
The characteristics of internal breast cancer in mass screening (ABS).
Journal of Experimental Medicine 39(3-4):109-16. December 1992.

T Morimoto, et al.
The quality of mass screening for breast cancer by physical examination.
Surgery Today 23(3):200-4. 1993.

L Nystrom et al.
Breast Cancer Screening with Mammography (letter).
Lancet 341(8859):1531, 1993.

G N Peters, V G Vogel, W P Evans.
The Texas breast screening project: Part 1.
Southern Medical Journal 86(4):385-90. April 1993.

M A Richards, et al.
Advanced breast cancer: use of resources and cost implications.
British Journal of Cancer 67 (4):856-60. April 1993.

A Rodgers
Breast cancer screening- the current position (letter)
British Medical Journal 302(6789):1401-1402, June 8th 1991

D.V.Schapira, et al.
Mammography Screening Credit Card and Compliance (Summary).
Cancer 70(2):509-512, 15 July 1992.

S Shapiro et al.
Periodic screening for breast cancer: the HIP project and its sequelae 1963-86
Baltimore:John Hopkins University Press, 1988

EA Sickles
The usefulness of computers in managing the operation of a mammographic screening process
American Journal of Roentgenology 155:755-761, October 1990

P Skrabanek.
Breast Cancer Screening with Mammography (letter).
Lancet 341(8859):1531, 1993.

A Stacey-Clear et al.
Breast cancer survival among women under age 50 - is mammography detrimental?
Lancet 340: 8826, 991-994. October 24th 1992.

N.Wald, C.Frost, H.Cuckle.
Breast Cancer Screening the current position (letter).
British Medical Journal 302, p845, 6 April 1991.

R Warren.
Team learning and breast cancer screening
Lancet 338:514, August 24 1991

Sensitometry

D.R.Bednarek, S.Rudin.
Comparison of Modified Bootstrap and Conventional Sensitometry in Medical Radiography.
Application of Optical Instrumentation in Medicine VIII.
Proceedings of Society of Photographic Optical Institute of Engineers 233:2-6 1980.

D.R.Bednarek, S.Rudin, R.Wong.
Evaluation of sensitomentry at Mammographic Energies.
Medical Physics 8:576, 1981.

A Burch, D Goodman et al.
Comparison of various film densitometers.
British Journal of Radiology 66(786):552-555, 1993.

Fundamentals of Radiographic Photography.
Volume I: Image Quality Control.
Kodak Limited. 1978
ISBN 0 901023 27 2

Fundamentals of Radiographic Photography.
Volume II: X-ray Recording Materials.
Kodak Limited. 1980
ISBN x 27 134958 0

Fundamentals of Radiographic Photography.
Volume III: Radiographic Quality.
Kodak Limited. 1982
ISBNx 27 134959 9

J Hale, P Bloch
Stepwedge sensitometry
Radiology 128:820-821, September 1978

J Hale and J W Thomas
Reciprocity law failure for a film-screen combination: Very long exposure times
Health Physics 45(3):780-782 September 1983

A G Haus
Sensitometric techniques in medical imaging
The Physics of Medical Imaging: Recording System Measurements and Techniques
American Institute of Physics 1989: 83-104

A G Haus and K Rossmann
X-ray sensitometer for screen/film combinations used in medical radiology
Radiology 94:673-678, 1970.

K.H.Thunthy, C.H.Boozer, R.Weinberg.
Sensitometric Evaluation of Rare-earth intensifying Screen Systems.
Oral Surgery Oral Medicine Oral Pathology 59(1): 102-6 January 1985.

C J Vyborny
H and D curves of screen-film systems: factors affecting their dependance on x-ray energy
Medical Physics 6:39-44, January-February 1979.

L.K.Wagner, G.T.Barnes, J.A.Bencomo, A.G.Haus.
An Examination of Errors in Characteristic Curve Measurement of Radiographic Screen/Film Systems.
Medical Physics 10:365-369, May/June 1983.

L.K.Wagner, A.G.Haus, G.T.Barnes, J.A.Bencomo, S.R.Amtey.
Comparison of Methods used to Measure the Characteristic Curves of Radiographic Screen/film systems.
Proceedings of Society of Photographic Optical Institute of Engineers 233:7-10 1980.

M.V.Yester, G.T.Barnes, M.K.King.
Kilovoltage Bootstrap Sensitometry.
Radiology 136:785-786 1980.

A.Yoshida, Y.Kiraki, Y.Ohtawa, T.Yamada, K.Hashimoto, K.Aono.
Modified Inverse Square Sensitometry for the Determination of the Characteristic Curve of Radiographic Screen/Film Systems.
Acta Medica Okayama 40:33-8, 1986.

Statistical analysis

JS Bulman, JF Osborne
Measuring diagnostic consistency
British Journal of Dentistry 116:377-381, May 28th 1989

HE Solberg, S Skrede and JP Blomhoff
Diagnosis of liver disease by laboratory results and discriminant analysis
Scandinavian Journal of Clinical Laboratory Investigation 35:713-721, 1975

HE Solberg
Discriminant analysis in clinical chemistry (editorial)
Scand. Journal of Clinical Laboratory Investigation 35:705-712, 1975

Treatment

R Calle, J Pillerson, P Schlienger, JR Vilcoq
Conservative Management of Operable Breast Cancer: Ten years experience at the Foundation ?
*Cancer.*2:2045-2053, 1978

JL Hayward, PJ Winter, D Tong, RD Rubens, JG Payne, MA Chaudary, F Habiollahi
A new combined approach to the conservative treatment of early breast cancer.
Surgery 95(3):270-274, 1984

B H Mason, I M Holdaway, N M Benlon, et al.
Detection of contralateral breast cancer by mammography in women with previous breast cancer and the impact of endocrine therapy .
New Zealand Medical Journal 106(949): 23-5. February 10th 1993.

SH Parker
Report of proceedings of 19th annual meeting of the Society of Cardiovascular and Interventional Radiology
San Diego 1994
RAD magazine p15 May 1994

G J G Rees.
Cancer treatment: deciding what we can afford.
British Medical Journal , 302. 6th April 1991.

Ultrasound

Zoe Anderson.
The Quality assurance of Ultrasonic Breast Scanners.
MSc Thesis, City University, November 1990.

J A Evans, RD Moore, S C Metcalfe and M Davis.
Routine Quality Assurance of high frequency ultrasound breast scanners in a screening context.
British Journal Of Radiology 66(787):614-618, 1993.

L Kaizer, EK Fishell, JW Hunt, FS Foster and NF Boyd
Ultrasonographically defined parenchymal patterns of the breast: relationship to mammographic patterns and other risk factors for breast cancer
British Journal of Radiology 61(722):118-124, February 1988.

Appendix A

List of names used in WHATIF.XLS, the computer model of the screening process and the cells in the spreadsheet to which they refer.

All_cancers D42
Assessed D11
Assessment_likelihoods G23:H24
Available D5
Biopsies D24
Biopsy likelihoods G32:H33
Ca_free E10
Ca_inc D10
Ca_inc_at_assessment D23
Ca_inc_biopsy D32
Ca_missed_at_assess D25
Ca_missed_at_biopsy D34
Ca_missed_at_screening D12
Cancers_detected D39
Cancers_found D33
Cancers_in_unscreened D37
Clear_inc J4
DNA G7
False_negatives D38
False_pos_at_screening D40
FHSA_no D4
Interval_cancers D37
mu J2
mu2 ..... K2
N H10:H11
N_a H23:H24
N_b H33:H34
noise G11:H11
noise_a G24:H24
noise_b G33:H33
Reg_errrors G5
Returned_letters G6
S G10:G11
S_a G23:G24
S_b G32:G33

Screened D6
Screening_likelihoods G10:H11
sigma J3
sigmaa J16
sigmab J29
signal G10:H10
signal_a G23:H23
signal_b G32:H32

	A	B	C
1	**Breast Screening Sensitivity Analysis**		
2	Lynn Martinez		
3			
4	No on FHSA register		
5	No available		
6	Screened		
7			
8	Cancer inc from data		
9			
10	inc of Cancer		
11	Assessed		
12	Missed Cancers (screening)		
13	tp	=signal S*Ca_inc	=B13*Screened
14	fp	=noise S*Ca_free	=B14*Screened
15	tn	=noise N*Ca_free	=B15*Screened
16	fn	=signal N*Ca_inc	=B16*Screened
17	p(tp)	=B13/(B13+B16)	=SUM(C13:C16)
18	p(fp)	=B14/(B14+B15)	
19	p(tn)	=B15/(B15+B14)	
20	p(fn)	=B16/(B16+B13)	
21			
22			
23	Ca inc in assessment population.		
24	Biopsy		
25	Missed Cancers (assessment)		
26			
27			
28			
29			
30			
31			
32	Ca inc in biopsy population		
33	Cancers found at biopsy		
34	Missed Cancers (biopsy)		
35			
36			
37	Cancers in unscreened population		
38	Cancers missed by screening		
39	Cancers Found by Screening		
40	False positives at Screening		
41			
42	All cancers		

	D
1	
2	
3	
4	100000
5	=FHSA_no-(FHSA_no*Reg_errors)
6	=Available-(Available*(Returned_letters+DNA))
7	
8	=(J11+K11)/L13
9	s
10	0.00775058275058275
11	=Screened*(signal S*Ca_inc+noise S*Ca_free)
12	=Screened*Ca_inc*signal N
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	s
23	=Ca_inc*signal S/(Ca_inc*signal S+Ca_free*noise S)
24	=Assessed*(signal_a S_a*Ca_Inc_at_assessment+G24*E23)
25	=Assessed*Ca_Inc_at_assessment*signal_a N_a
26	
27	
28	
29	
30	
31	s
32	=Ca_Inc_at_assessment*signal_a S_a/(Ca_Inc_at_assessment*signal_a S_a+E23*noise_a S_a)
33	=Biopsies*((signal_b S_b)*Ca_inc_biopsy+noise_b S_b*E32)
34	=Biopsies*Ca_inc_biopsy*(signal_b N_b)
35	
36	
37	=Ca_inc*(Available-Screened)
38	=SUM(Ca_missed_at_screening,Ca_missed_at_assess,Ca_missed_at_biopsy)
39	=Cancers_found
40	=Assessed-Cancers_found
41	
42	=SUM(D37:D39)

	E	F	G	H	I
1					Normal distribution
2					mu
3					sigma
4					inc
5		Registration errors	0.05		x
6		Returned letters	0.1		0.471669886032124
7		Did not attend/opted out	0.3		
8	=(J12+K12)/L13				
9	n	Screening likelihoods	S	N	**SCREENING**
10	=1-Ca_inc	signal	=T12	=U12	DATA
11		noise	=T13	=U13	s
12					n
13					
14					Normal distribution
15					
16					sigmaa
17					inc
18					x
19					0.609579139856017
20					
21					
22	n	Assessment likelihoods	S_a	N_a	**ASSESSMENT**
23	=1-Ca_Inc_at_assessment	signal_a	=T24	=U24	DATA
24		noise_a	=T25	=U25	s
25					n
26					Data inc
27					Normal Distribution
28					
29					sigmab
30					inc
31	n	Biopsy likelihoods	S_b	N_b	x
32	=1-Ca_inc_biopsy	signal_b	=T37	=U37	0.5
33		noise_b	=T38	=U38	
34					
35					**BIOPSY**
36					DATA
37					s
38					n
39					Data inc
40					
41					
42					

	J
1	1
2	0
3	0.316684522400902
4	=1-Ca_inc
5	y1(x)
6	=$B$4*EXP(-0.5*((I6-mu)/sigma)^2)/(sigma*(SQRT(2*PI()))
7	
8	
9	
10	S
11	1164
12	8928
13	
14	1
15	
16	0.358066032962379
17	=1-Ca_Inc_at_assessment
18	y1(x)
19	=$B$4*EXP(-0.5*((I19-mu)/sigmaa)^2)/(sigmaa*(SQRT(2*PI()))
20	
21	
22	
23	S
24	1164
25	388
26	=(J24+K24)/L26
27	1
28	
29	0.1
30	=1-K30
31	y1(x)
32	=$B$4*EXP(-0.5*((I32-mu)/sigmab)^2)/(sigmab*(SQRT(2*PI()))
33	
34	
35	
36	S
37	1143
38	0
39	=J37/L39
40	
41	
42	

	K	L
1	2	
2	1	
3		
4	=Ca_inc	(from LHS)
5	y2(x)	std 1
6	=$C$4*EXP(-0.5*((I6-mu2)/sigma)^2)/(sigma*(SQRT(2*PI()))))	=STANDARDIZE(I6,mu,sigma)
7		
8		
9		
10	N	
11	94	
12	179959	
13		=SUM(J11:K12)
14	2	
15		
16		
17	=Ca_Inc_at_assessment	
18	y2(x)	std 1
19	=$C$4*EXP(-0.5*((I19-mu2)/sigmaa)^2)/(sigmaa*(SQRT(2*PI()))))	=STANDARDIZE(I19,mu,sigmaa)
20		
21		
22		
23	N	
24	0	
25	8540	
26		=SUM(J24:K25)
27	2	
28		
29		
30	=D32	
31	y2(x)	std 1
32	=$C$4*EXP(-0.5*((I32-mu2)/sigmab)^2)/(sigmab*(SQRT(2*PI()))))	=STANDARDIZE(I32,mu,sigmab)
33		
34		
35		
36	N	
37	0	
38	388	
39		=SUM(J37:K38)
40		
41		
42		

	M	N
1	Best fit sigma	0.313549032080101
2	Best fit decision point	0.524077651146805
3	Intersection point	=((mu2+mu)/2)-(sigma*sigma*LN(Ca_inc/(1-Ca_inc))/(mu2+mu))
4		
5	std 2	area1=tn
6	=STANDARDIZE(I6,mu2,sigma)	=$J$4*NORMSDIST(L6)
7	DATA	=K12/L13
8		TN
9	NormDifference^2	=((N7-N6)/(N6+N7))^2
10	Intervals due to assessment	
11	0	
12		
13		
14	Best fit sigmaa	0.358066032962379
15	Best fit decision point	0.609579139856017
16	Intersection point	=((mu2+mu)/2)-(sigmaa*sigmaa*LN(K17/J17)/(mu2+mu))
17		
18	std 2	area1=tn
19	=STANDARDIZE(I19,mu2,sigmaa)	=(1-Ca_Inc_at_assessment)*NORMSDIST(L19)
20	DATA	=K25/L26
21		TN
22	NormDifference^2	=((N20-N19)/(N19+N20))^2
23		
24		
25		
26		
27	Best fit sigmab	0.1
28	Best fit decision point	0.5
29	Intersection point	=((mu2+mu)/2)-(sigma*sigma*LN(K30/J30)/(mu2+mu))
30		
31	std 2	area1=tn
32	=STANDARDIZE(I32,mu2,sigmab)	=$J$30*NORMSDIST(L32)
33	DATA	=K38/L39
34		TN
35	Difference^2	=((N33-N32)/(N32+N33))^2
36		
37		
38		
39		
40		
41		
42		

	O	P
1	=N1*0.1+N1	=N1-N1*0.1
2	=N2+N2*0.1	=N2-N2*0.1
3		
4		
5	area2+fn	tp
6	=$K$4*NORMSDIST(M6)	=$K$4-O6
7	=K11/L13	=J11/L13
8	FN	TP
9	=((O7-O6)/(O6+O7))^2	=((P7-P6)/(P6+P7))^2
10		
11		
12		
13		
14		
15		
16		
17		
18	area2+fn	tp
19	=Ca_Inc_at_assessment*NORMSDIST(M19)	=Ca_Inc_at_assessment-O19
20	=K24/L26	=J24/L26
21	FN	TP
22	=((O20-O19)/(O19+O20))^2	=((P20-P19)/(P19+P20))^2
23		
24		
25		
26		
27		
28		
29		
30		
31	area2+fn	tp
32	=$K$30*NORMSDIST(M32)	=$K$30-O32
33	=K37/L39	=J37/L39
34	FN	TP
35	=((O33-O32)/(O32+O33))^2	=((P33-P32)/(P32+P33))^2
36		
37		
38		
39		
40		
41		
42		

	Q	R	S	T
1			Screening ROC matrix	
2				
3			Outcome	S
4			s	=P6
5	fp		n	=Q6
6	=$J$4-N6			
7	=J12/L13		inc	s
8	FP	sum	Clears	n
9	=((Q7-Q6)/(Q6+Q7))^2	=SUM(N9:Q9)		
10			Likelihoods matrix solution	
11				S
12			s	=T4/U7
13			n	=T5/U8
14			Assessment ROC matrix	
15				S
16			s	=P19
17			n	=Q19
18	fp			
19	=(1-Ca_Inc_at_assessment)-N19		inc	s
20	=J25/L26		Clears	n
21	FP	sum		
22	=((Q20-Q19)/(Q19+Q20))^2	=SUM(N22:Q22)	Likelihood matrix solution	
23				S
24			s	=T16/U19
25			n	=T17/U20
26				
27			Biopsy ROC matrix	
28				S
29			s	=P32
30			n	=Q33
31	fp			
32	=$J$30-N32		inc	s
33	=J38/L39		Clears	n
34	FP	sum		
35	=((Q33-Q32)/(Q32+Q33))^2	=SUM(N35:Q35)	Likelihood matrix	
36				S
37			s	=T29/U32
38			n	=T30/U33
39				
40				
41				
42				

	U	V
1		
2		
3	N	
4	=O6	
5	=N6	
6		
7	=Ca_inc	
8	=Ca_free	
9		
10		
11	N	sum
12	=U4/U7	=SUM(T12:U12)
13	=U5/U8	=SUM(T13:U13)
14		
15	N	
16	=O19	
17	=N19	
18		
19	=Ca_Inc_at_assessment	
20	=1-Ca_Inc_at_assessment	
21		
22		
23	N	sum
24	=U16/U19	=SUM(T24:T24)
25	=U17/U20	=SUM(T25:U25)
26		
27		
28	N	
29	=O33	
30	=N32	
31		
32	=$D$32	
33	=1-U32	
34		
35		
36	N	
37	=U29/U32	
38	=U30/U33	
39		
40		
41		
42		

Appendix B

1st round									
Centre	a	b	c	d	e	f	g	h	total
No available/invited	31154	38570	45540	50255	39179	48165	15601	41275	309739
Screened	13029	28476	34350	32468	21281	32065	9788	18688	190145
% of invited	41.82	73.83	75.43	64.61	54.32	66.57	62.74	45.28	485
Assessed	912	1541	1551	1187	1049	1951	943	958	10092
% of invited	2.93	4.00	3.41	2.36	2.68	4.05	6.04	2.32	28
% of screened	7.00	5.41	4.52	3.66	4.93	6.08	9.63	5.13	46
c5	-46.39	3.79	-9.14	16.16	1.05	-7.59	-19.44	-10.87	-72
c5 estimate	0	4	0	16	1	0	0	0	21
Biopsied	186	195	260	255	123	246	113	153	1531
% of invited	0.60	0.51	0.57	0.51	0.31	0.51	0.72	0.37	4
% of screened	1.43	0.68	0.76	0.79	0.58	0.77	1.15	0.82	7
Malignant	109	161	197	213	104	190	80	110	1164
mal by biopsy	109	157	197	197	103	190	80	110	1143
% of invited	0.35	0.41	0.43	0.39	0.26	0.39	0.51	0.27	3
% of screened	0.84	0.55	0.57	0.61	0.48	0.59	0.82	0.59	5
Benign:Malig	0.20	0.24	0.26	0.30	0.19	0.25	0.14	0.27	2
ben	13	25	29	13	15	37	6	21	159
mal	66	104	111	44	77	151	44	79	676
biop	79	129	140	57	92	188	50	100	835
calc mal biopsy	155	157	206	197	103	198	99	121	1236
Intervals to early 91	3	3	11	45	6	2	0	3	73
Intervals 1st round	6	10	25	64	25	45	0	13	188
Missed=intervals/2	3	5	13	32	13	23	0	7	94
G(S/s)	109	161	197	213	104	190	80	110	1164
ass	109	161	197	213	104	190	80	110	1164
bio	109	157	197	197	103	190	80	110	1143
G(N/n)	12114	26930	32786	31249	20219	30091	8845	17723	179959
ass	726	1342	1291	916	925	1705	830	805	8540
bio	77	38	63	58	20	56	33	43	388
G(S/n)	803	1380	1354	974	945	1761	863	848	8928
ass	77	38	63	58	20	56	33	43	388
bio	0	0	0	0	0	0	0	0	0
G(N/s)	3	5	13	32	13	23	0	7	94
ass	0	0	0	0	0	0	0	0	0
bio	0	0	0	0	0	0	0	0	0
Screening									
Sensitivity	0.9754	0.9697	0.9397	0.8703	0.8909	0.8941	1.0000	0.9441	0.9253
Specificity	0.9378	0.9513	0.9603	0.9698	0.9553	0.9447	0.9111	0.9543	0.9527
Accuracy	0.9382	0.9514	0.9602	0.9690	0.9550	0.9444	0.9118	0.9543	0.9526
PPV	0.1195	0.1045	0.1270	0.1794	0.0991	0.0974	0.0848	0.1148	0.1153
NPV	0.9998	0.9998	0.9996	0.9990	0.9994	0.9993	1.0000	0.9996	0.9995
Assessment									
Sensitivity	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Specificity	0.9041	0.9725	0.9535	0.9405	0.9788	0.9682	0.9618	0.9493	0.9565
Accuracy	0.9156	0.9753	0.9594	0.9511	0.9809	0.9713	0.9650	0.9551	0.9616
PPV	0.5860	0.8090	0.7577	0.7860	0.8387	0.7724	0.7080	0.7190	0.7500
NPV	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Biopsy									
Sensitivity	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Specificity	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Accuracy	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
PPV	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
NPV	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

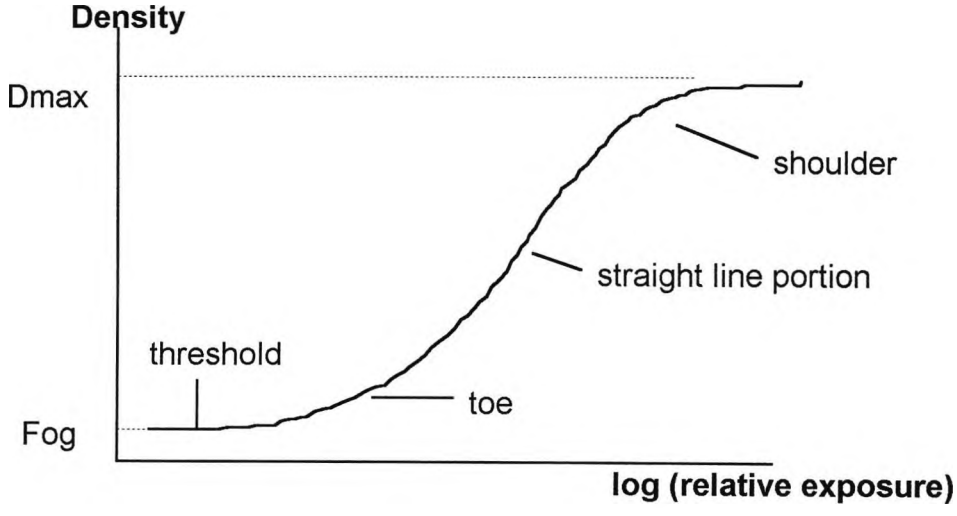
Appendix B

2nd round								
Centre	b	c	d	e	f	g	h	total
No available/invited	12566	21627	39494	10802	15927	7879	31341	139636
Self referrals	267	2268	7027	471	489	173	733	11428
Screened	8814	13661	25189	6542	10929	4961	12940	83036
% of invited	70.14	63.17	63.78	60.56	68.62	62.96	41.29	430.5221704
Assessed	426	437	677	299	394	317	395	2945
% of invited	3.39	2.02	1.71	2.77	2.47	4.02	1.26	17.65038279
% of screened	4.83	3.20	2.69	4.57	3.61	6.39	3.05	28.33773447
C5	15.23	-0.12	8.40	6.61	5.66	0.76	-43.75	-7.20709154
c5 estimate	15.00	0.00	8.00	7.00	6.00	1.00	0.00	37
Biopsied	27	67	107	16	44	23	75	359
% of invited	0.21	0.31	0.27	0.15	0.28	0.29	0.24	1.751180201
% of screened	0.31	0.49	0.42	0.24	0.40	0.46	0.58	2.911953162
Malignant	37	53	91	20	41	21	65	328
mal by biposy	22.00	53.00	83.00	13.00	35.00	20.00	65.00	291
% of invited	0.29	0.25	0.23	0.19	0.26	0.27	0.21	1.686426861
% of screened	0.42	0.39	0.36	0.31	0.38	0.42	0.50	2.775506992
Benign:Malig	0.24	0.26	0.30	0.19	0.25	0.14	0.23	2
ben	25	29	13	15	37	6	34	159
mal	104	111	44	77	151	44	145	676
biop	129	140	57	92	188	50	100	756
calc mal biop	22	53	83	13	35	20	109	335
Intervals to early 91	3	11	45	6	2	0	6	73
Intervals 2nd round	3	12	50	7	15	0	7	94
Missed=intervals/2	2	6	25	4	7	0	4	47
G(S/s)	37	53	91	20	41	21	65	328
ass	37	53	91	20	41	21	65	328
bio	22	53	83	13	35	20	65	291
G(N/n)	8386	13218	24487	6239	10528	4644	12541	80044
ass	384	370	562	276	344	293	320	2549
bio	5	14	24	3	9	3	10	68
G(S/n)	389	384	586	279	353	296	330	2617
ass	5	14	24	3	9	3	10	68
bio	0	0	0	0	0	0	0	0
G(N/s)	2	6	25	4	7	0	4	47
ass	0	0	0	0	0	0	0	0
bio	0	0	0	0	0	0	0	0
Screening								
Sensitivity	0.9576	0.8983	0.7848	0.8507	0.8464	1.0000	0.9482	0.8745
Specificity	0.9557	0.9718	0.9766	0.9572	0.9676	0.9401	0.9744	0.9683
Accuracy	0.9557	0.9715	0.9757	0.9568	0.9670	0.9403	0.9742	0.9679
PPV	0.0869	0.1213	0.1344	0.0669	0.1041	0.0662	0.1646	0.1114
NPV	0.9998	0.9995	0.9990	0.9994	0.9993	1.0000	0.9997	0.9994
Assessment								
Sensitivity	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Specificity	0.9871	0.9635	0.9590	0.9892	0.9745	0.9899	0.9697	0.9740
Accuracy	0.9883	0.9680	0.9645	0.9900	0.9772	0.9905	0.9747	0.9769
PPV	0.8810	0.7910	0.7913	0.8696	0.8200	0.8750	0.8667	0.8283
NPV	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Biopsy								
Sensitivity	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Specificity	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Accuracy	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
PPV	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
NPV	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Appendix C
Sensitometry

Sensitometry is a process where the response of a film to either light or X-ray exposure is measured. When the function of the film and processor are being considered in isolation, which we call *process* control, the film is exposed to a range of light levels using a device called a sensitometer. Usually, 21 areas of different exposure are produced by the sensitometer which, when developed, forms a grey scale step wedge on the film (example attached). The intensity of light producing each step is chosen so that each step is 0.15 units apart on the log relative exposure axis (two steps equals a doubling of intensity, one step multiplies the intensity of adjacent steps by 1.414). The colour of light emitted by the sensitometer must match the spectral sensitivity of the film so for X-ray films the choice of blue or green light is available. Mammography film is sensitive to blue light so that it is most sensitive to the spectrum emitted by rare earth screens.

The standard way to plot the results is to have the measured optical density on the y-axis and $\log_{10}$(relative exposure) on the x-axis. A typical curve looks like this:



Four parameters which describe this curve are routinely measured, fog (or base + fog), speed, slope and Dmax.

The **fog** (or sometimes base + fog) is simply the optical density of the film which has not received any direct exposure. Net densities are densities which have had the fog subtracted.

The threshold is the region where the film begins to show a response to exposure. As the exposure increases beyond the threshold the toe region occurs. In the toe region the gradient of the curve is beginning to increase and although images can be produced by exposures in this region, the final contrast will be low because the contrast amplification is low in this region.

The **slope or gradient** refers to the slope of the straight line portion of the film between the toe and the shoulder. The concept of the straight line portion is only approximate, the gradient of the curve actually changes continuously, however the approximation allows the above definition of gradient to be used in a practical situation. The convention is to use the points corresponding to a density of 0.25 + fog and 2 + fog to calculate the gradient, these points corresponding to the useful density range for diagnostic imaging, where the contrast amplification is greatest.

The **speed** is a comparative measure of the exposure required to produce a density of 1+fog. In order to produce some level of standardisation, film speeds are referred to a standard speed of 100, a film which is twice as fast as this (i.e. needs half of the exposure to produce the same blackening) is assigned a speed of 200.

Beyond the straight line portion the gradient of the film starts to fall, this is called the shoulder of the curve. The film will then reach **Dmax** which is the maximum density which can be recorded by the film (Dmax is not always achieved in sensitometry) and if the exposures are raised any further, a process known as solarisation begins to occur and the density then reduces as the exposure is increased. Solarisation has many uses but is not used in mammography.

In process control these parameters are measured and recorded every day on a control chart. Control limits are drawn on the chart which correspond to two standard deviations when the process is in control. If the parameters go outside of these limits, further investigations are made to find out the reason and if necessary, the processor is suspended from use.

It is also possible to perform sensitometry using X-rays as a source of exposure. The process is similar but the information produced will give a more accurate representation of how the film and screen will work in practice.

Sensitometry strip

Appendix D
Calculation of contrast

Using the equation

$$t = \frac{n_1 - n_2}{\sqrt{n_1 + n_2}}$$

for 95% confidence, must have t=1.96

$$1.96^2 = \frac{(n_1 - n_2)^2}{n_1 + n_2}$$

now if we have an intrinsic X-ray contrast of 1%,

$n_1 = 0.99\, n_2$

so the equation becomes

$$1.96^2 = \frac{(-0.01 \times n_2)^2}{1.99 \times n_2}$$

$$1.96^2 \times 1.99 n_2 = 0.01^2 \times n_2^2$$

$$\frac{1.96^2 \times 1.99}{0.01^2} = n_2$$

giving

n_2 = 71, 839

Similarly if the intrinsic contrast is 5%,

$n_1 = 0.95\, n_2$

and the calculation can be performed in the same manner as before.

Appendix E Statistical calculations

Minitab
Changes due to new tube
Screening
Expected counts are printed below observed counts

	Before	After	Total
Screen positive	171	189	360
	163.14	*196.86*	
Screen negative	3510	4253	7763
	3517.86	*4245.14*	
Total	3681	4442	8123
ChiSq =	0.379 +	0.314 +	
	0.018 +	0.015	= 0.725

df = 1

Minitab
Changes due to new tube - none expected because assessment is done on a different machine
Assessment
Expected counts are printed below observed counts

	Before	After	Total
Assessment positive	21	17	38
	18.57	*19.43*	
Assessment negative	150	162	312
	152.43	*159.57*	
Total	171	179	350
ChiSq =	0.319 +	0.305 +	
	0.039 +	0.037	= 0.700

df = 1

Minitab

Comparison of assessment at Southend and Romford
Expected counts are printed below observed counts

	Southend	Romford	Total
Clear at assessment	1705 *1702.66*	1346 *1348.34*	3051
Benign biopsy	56 *52.46*	38 *41.54*	94
Cancers found	190 *195.88*	161 *155.12*	351
Total number assessed	1951	1545	3496
ChiSq =	0.003 + 0.239 + 0.177 +	0.004 + 0.302 + 0.223 =	0.948

df = 2

The effect of processing on image quality. Straight line regression of processing parameters on the image quality index using Excel statistical analysis.

Regression Statistics						
Multiple R	0.984386286					
R Square	0.96901636					
Adjusted R Square	0.814098162					
Standard Error	0.663634166					
Observations	7					
Analysis of Variance						
	df	*Sum of Squares*	*Mean Square*	*F*	*Significance F*	
Regression	5	13.77387541	2.75477508	6.2550196	0.29421599	
Residual	1	0.440410306	0.44041031			
Total	6	14.21428571				
	Coefficients	*Standard Error*	*t Statistic*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	48.37974614	60.61697825	0.79812204	0.4552194	-721.82869	818.5882
Dmax	2.410507815	3.228896286	0.74654235	0.4835547	-38.616334	43.43735
Fog	-107.33534	22.01945298	-4.87456888	0.0027815	-387.11782	172.4471
Speed	-0.04399681	0.035443626	-1.24131788	0.260815	-0.4943488	0.406355
Gradient	1.066639498	11.30057527	0.09438807	0.9278741	-142.52017	144.6534
mAs	-0.57146979	0.348205939	-1.64118334	0.1518683	-4.9958268	3.852887