

Use of Software Tools to Implement Quality Control of Ultrasound Images in a Large Clinical Trial

William Stott

Thesis submitted for the degree of Doctor of Philosophy at University College London

2016

Department of Women's Cancer
EGA Institute for Women's Health
Faculty of Population Health Sciences
University College London

Declaration

I, William Stott, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Research Question

This thesis aims to answer the question as to whether software tools might be developed for automating the analysis of images used to measure ovaries in transvaginal sonography (TVS) exams. Such tools would allow the routine collection of independent and objective metrics at low cost and might be used to drive a programme of continuous Quality Improvement (QI) in TVS scanning. The tools will be assessed by processing images from thousands of TVS exams performed by the United Kingdom Collaborative Trial of Ovarian Cancer Screening (UKCTOCS).

Background

This research is important because TVS is core to any ovarian cancer (OC) screening strategy yet independent and objective quality control (QC) metrics for this procedure are not routinely obtained due to the high cost of manual image inspection. Improving the quality of TVS in the National Health Service (NHS) would assist in the early diagnosis of the disease and result in improved outcome for some women. Therefore, the research has clear translational potential for the >1.2 million scans performed annually by the NHS.

Research Findings

A study performed to process images from 1,000 TVS exams has shown the tool produces accurate and reliable QC metrics. A further study revealed that over half of these exams should have been classified as unsatisfactory as an expert review of the images showed that the sonographer had mistakenly measured a structure that was not an ovary. It also reported a correlation between such ovary visualisation and a novel metric (DCR) measured by the tools from the examination images.

Conclusion

The research results suggest both a need to improve the quality of TVS scanning and the viability of achieving this objective by introducing a QI programme driven by metrics gathered by software tools able to analyze the images used to measure ovaries.

Contents

Declaration.....	3
Abstract.....	5
Contents.....	7
List of Chapter Sections.....	9
List of Figures.....	21
List of Tables.....	25
List of Acronyms.....	27
Dedication.....	35
Acknowledgements.....	37
Introductory Chapters	
1 Problem Domain of the Thesis.....	45
2 Role of Technology in Quality of TVS Scanning.....	75
3 Managing the Quality of TVS Scanning.....	107
Tool Chapters	
4 Ultrasound Record Archive and Associated Tools.....	159
5 Tools for TVS Image Reviews.....	181
6 Image Processing Workflow Tool.....	201
Research Chapters	
7 Using the Image Workflow for TVS Quality Control.....	255
8 Audit of TVS Examinations by a Team of Experts.....	281
9 Automated Collection of TVS Quality Control Metrics.....	301
Conclusion Chapter	
10 Thesis Conclusion.....	321

Claims as to novelty and contribution to knowledge.....	335
---	-----

Appendices

A	Source of Data for Research Studies.....	341
B	Extracts from the UKCTOCS Protocol	343
C	UKCTOCS Scanning Report Form.....	347
D	UKCTOCS TVS Scan Classification Algorithm (flow diagram).....	349
E	Ultrasound Quality Assurance in UKCTOCS.....	351
F	Enhanced Anonymization of TVS Images.....	357
G	Comparison of Traditional and Cloud Datacentres.....	359
H	DICOM File Structure.....	365
I	Processor Types Used in Ultrasound Signal Processing.....	369
J	Review of Image Processing Tools.....	371
K	Data Quality.....	379

References

References.....	387
-----------------	-----

List of Chapter Sections

Chapter 1: Problem Domain of the Thesis

Introduction.....	45
Problem Statement.....	46
Importance of Quality Control in Ovarian Cancer Screening.....	46
Impact of Ovarian Cancer.....	46
Anatomy and Pathology of the Uterine Adnexa.....	47
Cancer in the Uterine Adnexa.....	50
Diagnostic Accuracy of Tests for Ovarian Cancer.....	50
Clinical Prediction of Ovarian Malignancy.....	53
Pathological Assessment of Ovarian Malignancy.....	55
Development of Ovarian Malignancy.....	59
UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS).....	62
Recruitment and Randomisation.....	62
Inclusion and Exclusion Criteria.....	63
Ultrasound Screening.....	63
TVS Examination.....	63
Images used for Studies.....	66
Completion of the Scan Report Form.....	67
Classification of TVS Examinations.....	68
Scanning Workflow.....	69
Statutory and Trial Regulations for UKCTOCS.....	71
Data Protection Act Compliance.....	71
Ethics approval.....	71
Summary.....	72

Chapter 2: Role of Technology in Quality of TVS Scanning

Introduction.....	75
Ultrasound Principles.....	76
Absorption and Reflection of Ultrasound.....	76
Resolution.....	76
Modes of Operation.....	77
Beam Focusing.....	78
Specification of Ultrasound Machine Resolution.....	79
Challenges of Image Interpretation.....	81
Safety.....	82
Review of Imaging Technologies used for Cancer Screening.....	83
Technical description of an ultrasound machine.....	86
Probes and cables.....	86
Front-end electronics.....	89
Data Compression.....	92
Digital Filtering.....	93
Changing the nature of the signal processing pipeline.....	93
Cine Memory.....	94
HistoScanning.....	95
Back-end Computer.....	95
Display.....	97
User Controls for Specific Ultrasound Functions.....	99
System Integration.....	102
Selection of Ultrasound Equipment.....	102
Machines used by UKCTOCS.....	103
Development of Ultrasound Technology and Future Trends.....	104
Summary.....	106

Chapter 3: Managing the Quality of TVS Scanning

Introduction.....	107
Concepts of Quality Assurance, Quality Control and Quality Improvement.....	108
Use of Statistics in Quality Improvement – a historical perspective.....	109
Quality Improvement using Six-Sigma Methodology.....	111
Measuring Process Capability.....	111
Removal of Special Variation.....	112
Methodology – DMAIC.....	113
Tool Selection.....	114
Accreditation and Training.....	115
Implementation Challenges and Criticisms.....	116
Approaches to Quality Improvement in Healthcare.....	117
Quality Improvement Studies.....	117
Quality Improvement Guidelines for UK Healthcare.....	119
Key Drivers of QI in UK.....	119
Key Organisations Providing QI Guidance in UK.....	120
Provision of QI Guidance in England.....	121
National Institute for Health and Care Excellence (NICE).....	121
National Quality Board (NQB)	122
Provision of QI Guidance by UK Healthcare Professional Bodies.....	123
Quality Assurance of Medical Ultrasound.....	124
Quality Assurance of Sonographers.....	124
Quality Assurance of Ultrasound Equipment.....	125
Technical Level Quality Checks.....	126
User Level Quality Checks.....	127
Quality Assurance of Ultrasound Procedures.....	129
General Considerations for the QA of Ultrasound Procedures...	130

Quality of Specific Ultrasound Procedures.....	131
Quality Assurance of TVS in Ovarian Cancer Screening Trials.....	138
Quality Assurance of TVS in PLCO.....	138
Quality Assurance of TVS in UKCTOCS.....	141
Published Papers.....	141
UKCTOCS Quality Assurance of TVS Scanning.....	142
Development of Standards and Accreditation.....	144
Fail-Safe Monitoring of Scan Report Forms.....	145
Monitoring of Abnormal Scans and Image Reviews.....	145
Sonographer QA.....	145
Equipment QA and Role of IT Systems.....	147
Scanning Process QA.....	148
Summary.....	153

Chapter 4: Ultrasound Record Archive and Associated Tools

Introduction.....	159
History of URA.....	160
Requirements Analysis.....	161
Integration with the UKCTOCS ultrasound scanning workflow.....	161
Processing disks from trial centres.....	161
Validation against data in TMS.....	161
Manual correction.....	161
Anonymization.....	162
Image Storage.....	162
Image Export.....	162
Non-functional Requirements.....	162

Design and Implementation.....	163
Discussion.....	164
Use of osImageManager for importing Trial Images.....	165
Processing and Storage of Images from Trial Centres.....	166
Data Stored in URA.....	167
Use of URA for Research – osImageManager.....	168
Further Development – URAImageExport.....	169
Requirements Analysis.....	169
Design and Implementation.....	170
Applications for URAImageExport.....	170
Future Plans.....	171
Proposed Institute for Women’s Health Biobank.....	172
Biobank Management System (BMS).....	172
Study Data Management System (SDMS).....	173
Image Management System (IMS).....	173
Trial Management System (TMS).....	174
Lab Information Management System (LIMS).....	175
Application: Image Collector.....	175
Application: Image Curator.....	176
Web Browser: Trial Centre Collaborator.....	176
Web Browser: Research Collaborator.....	176
Proposed open-source project.....	177
Summary.....	179

Chapter 5: Tools for TVS Image Reviews

Introduction.....	181
Spreadsheet used to confirm ovary visualisation.....	182
Requirement Analysis.....	182
Design.....	183
Implementation.....	184
Discussion.....	184
Application used to conduct an audit of TVS Exams.....	185
Requirement Analysis.....	185
Design.....	188
User Interface Layer.....	189
Business Logic Layer.....	190
Data Layer.....	190
Implementation.....	191
Discussion.....	191
General application for image annotation during reviews.....	192
Requirement Analysis.....	192
Design.....	194
Data Layer.....	195
Business Logic Layer.....	196
User Interface Layer.....	197
Implementation.....	198
Discussion.....	199
Summary.....	200

Chapter 6: Image Processing Workflow Tool

Introduction.....	201
Requirements Analysis.....	202
Problem Scope.....	202
Requirements for an Image Workflow.....	203
Definition of Image Collections.....	203
Processing.....	204
Support for Imager Processing Transactions.....	204
Common Problems of Research Involving Images.....	212
Review of Existing Work.....	214
Workflow Automation.....	215
Componentisation of Software for Image Processing.....	216
Problems of C++ binary level compatibility.....	217
COM Solution.....	219
OpenCV Image Processing Library.....	228
Design of a Workflow for Processing Images.....	228
Key Design Decisions.....	229
Design Overview.....	230
MxPlugin Architecture.....	231
MxPlugin Image Processing Components.....	236
Workflow Processing Software.....	238
Design of the Command-line User Interface – ICApps.....	242
Domain Specific Language Design.....	245
Design of the Graphical User Interface – ImageExpter.....	246
Design of Components for Processing UKCTOCS Images.....	247
MxPluginForImages component – OpenCV.....	248
MxPluginForImages components – UKCTOCS.....	248

Discussion of use of the workflow in research.....	249
Implementing Quality Control for UKCTOCS Images.....	249
Future Work.....	250
Further Development.....	250
Summary.....	252

Chapter 7: Using the Image Workflow for TVS Quality Control

Introduction.....	255
Background and Objectives.....	256
Methods.....	257
Identifying the Images Used to Measure the Ovaries.....	258
Image Workflow Processing Requirement.....	263
Finding the Calliper Marks.....	263
Finding the Length of Calliper Lines.....	264
Finding the D1, D2 Angle.....	264
Finding the Format of the Image.....	264
Design of Processing Components.....	265
Algorithm for Finding Calliper Lines and their Properties.....	265
Algorithm for Classifying Image and TVS Exam Formats.....	267
Implementation of Workflow Processing.....	269
Statistical Analysis.....	269
Identification of Images used to Measure Ovaries.....	269
Repeatability of Results.....	270
Manual Review.....	270
Agreement of Ovary Volumes: Image vs TMS Values.....	271
Calculation of Quality Control Metrics.....	272

Results.....	272
Identification of Images used to Measure Ovaries.....	273
Repeatability of Results.....	273
Manual Review.....	274
Agreement of Ovary Volumes: Image vs TMS values.....	275
Calculation of Quality Control Metrics.....	276
Discussion.....	277
Identification of the Exact Images Used for Measurement.....	278
Capability of the Workflow Software as a Measurement Device.....	278
Limitation of Study.....	279
Quality Control metrics Applied to UKCTOCS TVS Scanning.....	279
Summary.....	280
Chapter 8: Audit of TVS Examinations by a Team of Experts	
Introduction.....	281
Methods.....	283
Statistical analysis.....	286
Kappa Statistic.....	286
Results.....	288
Volunteer Variation.....	291
Observer variation between auditors.....	292
Resources required by the audit.....	292
Discussion.....	293
Factors causing variation in VR-Both amongst sonographers.....	293
Observer variation between the expert reviewers.....	294
Results compared with other work.....	295
Planning and execution of future studies.....	295

Limitations of the Study.....	296
Future Work.....	296
Summary.....	298

Chapter 9: Automated Collection of TVS Quality Control Metrics

Introduction.....	301
Background and Objectives.....	303
Methods.....	303
Expert Review.....	303
Requirements.....	304
Design and Implementation.....	304
Quality Control metrics obtained by software analysis of the images...	304
Requirements Analysis.....	304
Design and Implementation.....	306
Statistical Analysis.....	306
Results.....	308
Discussion.....	310
Visualisation Rate Obtained by Expert Review.....	311
Visualisation Rate Obtained by the Software Tool.....	313
Other Work.....	314
Limitations of Study.....	314
Future Work.....	314
Summary.....	316

Chapter 10: Thesis Conclusion

Introduction.....	321
Answer to the research question.....	322
Delivery of High Quality TVS Scanning.....	322
Audit.....	322
Quality Plan.....	323
Quality Governance Framework.....	324
TVS scanning used as a screen for ovarian cancer.....	325
Quality Control of Ultrasound Equipment.....	326
Training and Accreditation of Sonographers.....	327
Provision of Information Technology Systems.....	327
Collection of QC Metrics.....	328
Ability to capture images of an ovary.....	328
Placement of calliper marks.....	329
Dimensions Assigned Correctly (DAC).....	329
Static Image Quality.....	330
Conclusion.....	331
Impact of the thesis project.....	333
Enhancing the value of the URA.....	333
Useful tools for other researchers.....	333
Future Work.....	333
Open-source projects.....	333
Production version of the ImageExpter.....	334
Segmentation of ovaries from TVS images.....	334
Classifier to discriminate between images containing ovaries	334
Summary.....	337

List of Figures

1-1	Main Components of the Uterine Adnexa.....	48
1-2	ROC for optimisation of test sensitivity and specificity.....	52
1-3	Static image of a section of the uterus	65
1-4	Measurement of ovary dimensions.....	66
1-5	UKCTOCS Workflow for performing a TVS Scan.....	70
2-1	Resolution Integral Calculation.....	80
2-2	Medical Ultrasound System Block Diagram.....	86
2-3	Main Components of an Ultrasound Probe.....	87
2-4	Stepped and Phase Array Signal Formation.....	88
2-5	Comparison of Analogue and Digital Beamforming.....	91
2-6	Modern Ultrasound Signal Processing.....	94
2-7	Static Ultrasound Image Captured during a TVS Examination.....	99
2-8	Medison Accuvix XQ User Interface.....	101
2-9	Medison Accuvix XQ Ultrasound Machine.....	103
3-1	Control Chart for Variables.....	110
3-2	Six-Sigma – Zero Defects.....	113
3-3	Reverberation pattern produced by operation in free air.....	129
3-4	Nuchal Translucency Quality Control Reference Curve	135
3-5	Annual Visualisation Rates in Trial Centres.....	150
4-1	UKCTOCS Ultrasound Data Processing System.....	163
4-2	URA Main Window.....	164
4-3	URA File Import Dialog.....	165
4-4	UKCTOCS TVS Exams stored in the URA	167
4-5	Distribution of TVS Exams after 1/1/08 by Trial Centre.....	168
4-6	URA File Export Dialog.....	169
4-7	Format of URAIImageExport input file.....	170
4-8	URAIImageExport main window.....	171
4-9	Proposed Architecture for Institute for Women’s Health Biobank.....	173

5-1	Spreadsheet used for Image Review.....	184
5-2	osImageManager adapted for Image Review.....	189
5-3	Image Review Dialog Boxes.....	190
5-4	Differential Analysis of Ovary Boundary.....	194
5-5	Part of the body of an IDF file.....	196
5-6	ImageCurator User Interface.....	198
6-1	Environment for specialisation in research involving image analysis ...	213
6-2	COM object with a number of interfaces.....	225
6-3	MxPluginManagerGUI – Application for Managing MxPlugins.....	234
6-4	Workflow File Structure.....	239
6-5	Image Processing Collection File.....	241
6-6	ICApps System Overview.....	243
6-7	Language Workbench for creation of IPF files.....	246
6-8	Optimisation of algorithm and parameter selection.....	251
7-1	Example of a TSLS-RO image containing right ovary.....	256
7-2	Defect in the calculation of dimensions.....	259
7-3	Optimization of matching tolerance for dimensions.....	268
7-4	Success of identifying the exact images used for measurements.....	274
7-5	Scatter chart showing correlation of ovary volumes.....	275
7-6	Bland-Altman plot showing agreement of ovary volumes.....	276
8-1	Partitioning of sonographers’ TVS exams to the reviewers.....	284
8-2	Accuracy and Precision of marksmen shooting at a target.....	287
8-3	Variation in VR-Both Confirmed by Individual auditors.....	290
8-4	Confirmation of sonographer VR by the team of 9 auditors.....	290
8-5	Volunteer variation by trial centre; age, time last period and BMI.....	291
8-6	Volunteer variation by trial centre; hysterectomy, HRT, OCP usage.....	292
9-1	Correct measurement of an ovary.....	312
9-2	Measurement of bowel instead of ovary.....	312

A-1	Selection of Volunteers for TVS Scans in UKCTOCS.....	341
A-2	Source of TVS scan images used for thesis research studies.....	342
B-1	Ultrasound Algorithm (Figure 3 in UKCTOCS Protocol).....	345
C-1	Front Page of UKCTOCS Scanning Report.....	347
C-2	Rear Page of UKCTOCS Scanning Report.....	348
D-1	UKCTOCS TVS Scan Classification Algorithm (flow diagram).....	349
F-1	TVS static image with personal identifiable information removed.....	358
G-1	Cloud Datacentre.....	359
H-1	Contents of DICOM Header from UKCTOCS Image.....	367
J-1	Organisation of NAMIC activities.....	372

List of Tables

1-1	Test outcome results presented in terms of a 2 X 2 Table	51
4-1	Examples of Open Source Biobanks.....	177
7-1	UKCTOCS TVS Static Image Formats.....	261
7-2	UKCTOCS TVS Exam Image Formats.....	262
7-3	Novel quality control metrics for TVS exams.....	272
7-4	Results of calculating the DAC, D2O, D1L and DCR metrics.....	277
8-1	Observer variation between pairs of expert reviewers.....	293
9-1	DCR Metric Definition.....	305
9-2	Truth tables giving various definitions of Visualisation Rates (VR).....	307
9-3	Visual Rates (VR) from Expert Review by VR definition.....	308
9-4	Visualisation (VR) from Expert Review by exam subsets.....	309
9-5	Contingency table: DCR vs. Expert Review.....	310
9-6	Reports of ovary visualisation and VR definitions used.....	311

List of Acronyms

3D	Three Dimensional
AAA	Abdominal Aortic Aneurysm
ABF	Analogue Beam Forming
ACID	Atomic, Consistent, Isolated, Durable
ACMG	American College of Medical Genetics
ACRIN	American College of Radiology Imaging Network
ADC	Analogue to Digital Converter
API	Application Programming Interface
ASIC	Application Specific Integrated Circuit
BA	Bland Altman
BBOT	Brenner Borderline Ovarian Tumour
BLOB	Binary Large Object
BMP	Bitmap
BMS	Biobank Management System
BMUS	British Medical Ultrasound Society
BRCA1	breast cancer type 1 susceptibility protein
BRCA2	breast cancer type 2 susceptibility protein
BSP	Breast Screening Programme
CA-125	Cancer Antigen 125
CASE	Consortium for the Accreditation of Sonographic Education
CC	Coordinating Centre
CCBOT	Clear Cell Borderline Ovarian Tumour
CC-US	Coordinating Centre Ultrasound Scanning
CD	Compact Disk
CEO	Chief Operating Officer
CI	Confidence Interval
COM	Component Object Model
COTS	Commercial Off The Shelf

CPU	Central Processing Unit
CQC	Care Quality Commission
CRL	Crown Rump Length
CSV	Comma Separated Variable
CT	Computed Axial Tomography
CUI	Command-line User Interface
D1L	Dimension One Length
D2O	Dimension Two Degrees
DAC	Dimensions Assigned Correctly
DAC	Digital to Analogue Converter
DBF	Digital Beam Forming
DBMS	Database Management System
DCR	Dimensions in Correct Range
DEMUX	De-multiplexor
DICOM	Digital Imaging and Communications in Medicine
DID	Diagnostic Imaging Dataset
DLL	Dynamic Link Library
DMIAC	Define, Measure, Analyse, Improve, Control
DOE	Design of Experiments
DPA	Data Protection Act
DPMO	Defective Parts per Million Opportunities
DQASS	Down's syndrome screening Quality Assurance Support Service
DSL	Domain Specific Language
DSP	Digital Signal Processor
DTW	Dynamic Time Warping
EBOT	Endometriod Borderline Ovarian Tumour
EDC	Electronic Data Capture
EU	European Union
EXE	Executable

FASP	Fetal Anomaly Screening Programme
FIFO	First In First Out
FIGO	International Federation of Gynaecology and Obstetrics
FIPS	Federal Information Processing Standard
FMF	Fetal Medical Foundation
FN	False Negative
FP	False Positive
FPGA	Field Programmable Gate Array
FSDR	First Scan Detection Rate
GMC	General Medical Council
GPU	Graphic Processing Unit
GUI	Graphical User Interface
GUID	Global Unique IDentifier
HCPC	Health and Care Professions Council
HGSC	High Grade Serious Carcinoma
HIS	Hospital Information System
HPC	High Performance Cluster
HQIP	Healthcare Quality Improvement Partnership
HRT	Hormone Replacement Therapy
HV	High voltage
IBM	International Business Machines
ICD	International Classification of Diseases
IDF	Image Dataset File
IDL	Interface Definition Language
IEF	Image Experiment File
IJF	Image Job File
IMS	Image Management System
IOTA	International Ovarian Tumour Analysis
IPC	Image Processing Collection

IPF	Image Processing File
IPP	Intel Performance Primitives
IT	Information Technology
JIST	Java Image Science Toolkit
LIMS	Laboratory Management System
LNA	Low Noise Amplifier
LO	Left Ovary
LONI	Laboratory of Neuro Imaging
LS	Longitudinal Section
LS-LO	Longitudinal Section of Left Ovary
LS-RO	Longitudinal Section of Right Ovary
MBOT	Mucinous Borderline Ovarian Tumour
MDI	Multiple Document Interface
MFC	Microsoft Framework Classes
MI	Mechanical Index
MMS	Multimodal Screening
MO	Magneto-Optical
MoM	Multiples of Median
MOOP	Manual of Operations and Procedures
MPE	Medical Physics Expert
MRI	Magnetic Resonance Imaging
MUSIQ	Model for Understanding Success In Quality
MUX	Multiplexor
NHS	National Health Service (UK)
NHSLA	National Health Service Litigation Authority
NICE	National Institute for Health Care Excellence
NIPE	Neonatal and Infant Physical Examination
NLS	National Lead Sonographer
NLST	National Lung Screening Trial

NQB	National Quality Board
NSC	National Screening Committee
NT	Nuchal Translucency
OC	Ovarian Cancer
OCP	Oral Contraceptive Pill
OGE	Oracle Grid Engine
OO	Object Oriented
PACS	Picture Archiving and Communication System
PC	Personal Computer
PDCA	Plan, Do, Check, Act
PDF	Portable Document Format
PEF	Image Project File
PET	Positron Emission Tomography
PLCO	Prostrate, Lung, Colorectal, Ovarian
PPV	Positive Predictive Value
PSX	Processing Steps XML
QA	Quality Assurance
QACC	Quality Assurance Coordinating Centre
QC	Quality Control
QI	Quality Improvement
QUADIS	QUality Assessments of DIagnostic Accuracy Studies
RAM	Random Access Memory
RCM	Royal College of Midwives
RCOG	Royal College of Obstetricians and Gynaecologists
RCR	Royal College of Radiologists
RF	Radio Frequency
RMI	Risk of Malignancy Index
RO	Right Ovary
ROC	Receiver Operating Curve

ROC	Risk of Cancer
RPC	Remote Procedure Call
RQIA	Regulation and Quality Improvement Authority
Rx	Receive
SBOT	Serious Borderline Ovarian Tumour
SC	Screening Centre
SCoR	Society and College of Radiologists
SCSOCS	Shizuoka Cohort Study on Ovarian Cancer Screening
SDMS	Study Data Management System
SMBOT	Seromucinous Borderline Ovarian Tumour
SNR	Signal to Noise Ratio
SOAP	Simple Object Access Protocol
SOP	Standard Operating Procedure
SPL	Spatial Pulse Length
SPV	Sparse Pixel Vectorization
SQC	Statistical Quality Control
SQL	Structured Query Language
SQUIRE	Standards for Quality Improvement Reporting Excellence
SSS	Screening Support Sonographers
TAS	TransAbdominal Scan
TC	Trial Centre
TGC	Time Gain Compensator
TI	Thermal Index
TMM	Tissue Mimicking Material
TMS	Trial Management System
TN	True Negative
TP	True Positive
TPM	Transaction Processing Monitor
TQM	Total Quality Management

TS	Transverse Section
TS-LO	Transverse Section of Left Ovary
TSLs	Transverse Section and Longitudinal Section
TSLs-Both	Transverse Section and Longitudinal Section of Both left and right ovary
TSLs-Left	Transverse Section and Longitudinal Section of Left ovary
TSLs-LO	Transverse Section, Longitudinal Section of Left Ovary
TSLs-None	Transverse Section and Longitudinal Section of neither left or right ovary
TSLs-Right	Transverse Section and Longitudinal Section of Right ovary
TSLs-RO	Transverse Section, Longitudinal Section of Right Ovary
TS-RO	Transverse Section of Right Ovary
TVS	Transvaginal Ultrasound Scan
Tx	Transmit
UCL	University College London
UCLH	University College London Hospital
UK	United Kingdom
UKCTOCS	United Kingdom Collaborative Trial of Ovarian Cancer Screening
URA	Ultrasound Record Archive
USC	Ultrasound Subcommittee
USS	Ultrasound Screening
VGA	Variable Gain Amplifier
VMT	Virtual Method Table
VR	Visualisation Rate
VR-Both	Visualisation Rate – Both Ovaries seen
VRN	Volunteer Reference Number
WHO	World Health Organisation
WWF	Windows Workflow Foundation
XML	Extended Markup File

Dedication

In memory of Kathleen Stott and Joan Brearley who both died of ovarian cancer

Acknowledgements

The completion of a PhD is a rite of passage, so I must give my sincere gratitude to Dr. Chris Jones and Prof. Usha Menon for the help they have provided as supervisors to prepare me for making this transition into the next phase of my career. In addition there have been many others who have given particular support, encouragement and practical assistance during this journey. They know individually what they have done and the contribution they have made, so I will simply list their names in alphabetical order:

Kim Acland, Omar Al Okashi, Nazar Asmo, Simon & Ann Battensby, Oleg Blyuss, Peter Brocklehurst, Gabriel Brostow, Carol Brunell, Matthew Burnell, Stuart Campbell, John Dodgson, Anthea Ferguson, Gwen Fletcher, Kathy Ford, Angelo Franchini, Alex Gentry-Maharaj, Sioban Gupta, Allan Hackshaw, Jane Hayward, Anna Hyde, Dheyaa Ibrahim, Ian Jacobs, Tindie Kalsi, Robert Liston, Martina Markoss, Angela Poulter, Rani Ranger, Andy Ryan, Mourad Seif, Aarti Sharma, Gill Turner, Yuchen Wang, Alexy Zaikin

I must also acknowledge everyone who has helped me from my early teachers to college lecturers as well as the people who I've never met, but nevertheless have influenced my thinking through their publications or presentations. Furthermore there are friends, colleagues and family not mentioned by name yet still deserve acknowledgement – you know who you are. However, I must give specific thanks to my partner Mark Brearley not just for his fantastic support for my work at UCL over the last decade, but also for the perspective he provided when I was struggling to complete the concluding chapter. He considered the issues as a senior manager of a large global business and gave me the confidence to address the wider problems of quality management and organisational change. Finally acknowledgement needs to be given all the people involved in the trial, particularly the 48,250 women who kindly agreed to participate in UKCTOCS and year after year submitted themselves to TVS examination in the hope of making a worthwhile contribution to ovarian cancer research. All the researchers who use the information they have so generously provided must ensure it is put to good use. I have certainly attempted to do so.

Material and Assistance Provided by Others

In respect of the declaration made at the start of the thesis the following specific material and assistance provided by others must be credited in addition to the assistance from others acknowledged in the previous section.

Derived Artwork

The following artwork was created by Terry Castellani and derived from information given to him from various sources including those listed:

- Figure 1-1: Main Components of the Uterine Adnexa
 - H.Kenneth Walker, W.Dallas Hall, and J.Willis Hurst, *Clinical Methods: The History, Physical and Laboratory Examinations*, 3rd ed. (Butterworth-Heinemann Ltd, 1990).
 - Sanders, Roger C., and Tom C. Winter. 2006. *Clinical Sonography: A Practical Guide*. 4th Revised edition. Lippincott Williams and Wilkins.
 - Zuletzt geändert. 2016. 'Black and White Illustration of Naked Woman Showing Uterus, Fallopian...' *Getty Images*. Accessed June 24.
<http://www.gettyimages.ch/detail/illustration/black-and-white-illustration-of-naked-woman-lizenfreie-illustration/84289978>.
- Figure 2-2: Medical Ultrasound System Block Diagram
 - Brunner, Eberhard. 'Ultrasound Front-End IC Considerations - UltrasoundFrontend.pdf'. *Ultrasound System Considerations and Their Impact on Front- End Components*, 2002. <http://www.analog.com>
- Figure 2-3: Main Components of an Ultrasound Probe
 - MBA, Vivien Gibbs DCR DNM DMU MA, David Cole DCR DMU MSc, and Antonio Sassano Bsc MSc. 2009. *Ultrasound Physics and Technology: How, Why and When*, 1e. 1 edition. Edinburgh ; New York: Churchill Livingstone.
 - Analog Devices: *Analog Dialogue: How Ultrasound System Considerations Influence Front-End Component Choice*. 2015. Accessed December 22.
<http://www.analog.com>.
- Figure 2-4: Stepped and Phase Array Signal Formation
 - MBA, Vivien Gibbs DCR DNM DMU MA, David Cole DCR DMU MSc, and Antonio Sassano Bsc MSc. 2009. *Ultrasound Physics and Technology: How, Why and When*, 1e. 1 edition. Edinburgh ; New York: Churchill Livingstone.
- Figure 2-5: Comparison of Analogue and Digital Beamforming
 - 'The Future of Beamforming in Ultrasound | Medical Electronics Design'. 2015. Accessed December 23. <http://www.medicalelectronicsdesign.com>.
 - Kwon, Jiwon, Jae Hee Song, Sua Bae, Tai-kyoung Song, and Yangmo Yoo. 2012. 'An Effective Beamforming Algorithm for a GPU-Based Ultrasound Imaging

System'. In Ultrasonics Symposium (IUS), 2012 IEEE International, 619–22.
doi:10.1109/ULTSYM.2012.0154.

- **Figure 2-6: Modern Ultrasound Signal Processing**
 - Güracar, Ismayil. 2016. 'Experiences Porting Real Time Signal Processing Pipeline CUDA Kernels to Kepler and Windows 8'. Accessed June 24. <http://on-demand.gputechconf.com/gtc/2014/presentations/S4148-rt-signal-processing-pipeline-cuda-kernels-kepler-win8.pdf>.
- **Figure 3-1: Control Chart For Variables**
 - Krishnaiah, K. 2014. Applied Statistic Quality Control and Improvement. 1 edition. PHI Learning.
- **Figure 3-2: Six Sigma – Zero Defects**
 - Pyzdek, Thomas, and Paul Keller. 2014. The Six Sigma Handbook, Fourth Edition. 4 edition. New York: McGraw-Hill Education.
 - Krishnaiah, K. 2014. Applied Statistic Quality Control and Improvement. 1 edition. PHI Learning.
- **Figure 3-4: Measurement of Bias in NT and CRL measurement**
 - 'Fetal Anomaly Screening: Ultrasound Practitioners Handbook - Publications - GOV.UK'. 2016. Accessed March 21.
<https://www.gov.uk/government/publications/fetal-anomaly-screening-ultrasound-practitioners-handbook>.
- **Figure 8-2: Accuracy and Precision of marksmen shooting at a target**
 - 'CHEMISTREE11: October 2010'. 2016. Accessed June 15.
http://chemistree11.blogspot.com/2010_10_01_archive.html.
 - Viera, A. J., and J. M. Garrett. 2005. 'Understanding Interobserver Agreement: The Kappa Statistic'. Family Medicine 37 (5): 360–63
- **Figure G-1: Cloud Datacentre**
 - The Economist. 2015. 'The Cheap, Convenient Cloud', April 18.
<http://www.economist.com/news/business/21648685-cloud-computing-prices-keep-falling-whole-it-business-will-change-cheap-convenient>.

Reproduction of copyright material

The author did not create the following material, so permission has been obtained from the copyright holder for its reproduction in the thesis and credit is given as follows:

- Figure 1-1: Main Components of the Uterine Adnexa © Terry Castellani
- Figure 1-2: ROC for optimisation of test sensitivity and specificity © Terry Castellani
- Figure 1-3: Static image of a section of the uterus with endometrial. © UCL
- Figure 1-4: Measurement of ovary dimensions. © UCL
- Figure 2-1: Pye, S. D., W. Ellis, and T. MacGillivray. 2004. 'Medical Ultrasound: A New Metric of Performance for Grey-Scale Imaging'. *Journal of Physics: Conference Series* 1 (1): 187. doi:10.1088/1742-6596/1/1/041. © IOP Publishing. Reproduced with permission. All rights reserved
- Figure 2-2: Medical Ultrasound System Block Diagram © Terry Castellani
- Figure 2-3: Main Components of an Ultrasound Probe © Terry Castellani
- Figure 2-4: Stepped and Phase Array Signal Formation © Terry Castellani
- Figure 2-5: Comparison of Analogue and Digital Beamforming © Terry Castellani
- Figure 2-6: Modern Ultrasound System © Terry Castellani
- Figure 2-7: Static Ultrasound Image Captured during a TVS Examination © UCL
- Figure 2-8: Medison Accuvix XQ User Interface © Samsung Medison Co., LTD
- Figure 2-9: Medison Accuvix XQ Ultrasound Machine © Samsung Medison Co., LTD
- Figure 3-1: Control Chart For Variables © Terry Castellani
- Figure 3-2: Six Sigma – Zero Defects © Terry Castellani
- Figure 3-3: Reverberation pattern produced by operation in free air. © Steve Pye
- Figure 3-4: Measurement of Bias in NT and CRL measurement © Terry Castellani
- Figure 3-5: Annual Visualisation Rates in Trial Centres. © 2015 Sharma, A., M. Burnell, A. Gentry-Maharaj, S. Campbell, N. N. Amso, M. W. Seif, G. Fletcher, et al. 2016. 'Quality Assurance and Its Impact on Ovarian Visualization Rates in the Multicenter United Kingdom Collaborative Trial of Ovarian Cancer Screening (UKCTOCS)'. *Ultrasound in Obstetrics & Gynecology*: 47 (2): 228–35. doi:10.1002/uog.14929. Published by John Wiley & Sons Ltd on behalf of International Society of Ultrasound in Obstetrics and Gynecology.
- Figure 4-2: URA Main Window – ultrasound image © UCL
- Figure 5-2: osImageManager adapted for Image Review – ultrasound image © UCL
- Figure 5-4: Differential Analysis of Ovary Boundary – ultrasound image © UCL
- Figure 5-6: ImageCurator User Interface – ultrasound image © UCL
- Figure 7-1: Example of a TSLS-RO image containing right ovary © UCL
- Figure 7-2: Defect in the calculation of dimensions – ultrasound image © UCL
- Figure 9-1: Correct measurement of an ovary © UCL
- Figure 9-2: Measurement of bowel instead of ovary © UCL

- Figure B-1: Figure 3 from UKCTOCS protocol © UCL
- Figure C-1: Front Page of UKCTOCS Scanning Report © UCL
- Figure C-2: Rear Page of UKCTOCS Scanning Report © UCL
- Figure F-1: TVS static image with personal identifiable information removed © UCL
- Figure J-1: Organization of NAMIC activities © Harvard Medical School
- Appendix B: Extracts from the UKCTOCS protocol © UCL
- Appendix E: Ultrasound Quality Assurance in UKCTOCS © UCL

Sources of data and analysis

- Image assessments for the studies described in Chapters 7, 8 and 9 were performed by various experts in gynaecological scanning.
- Data related to TVS Exams stored in the Trial Management System - Andy Ryan, UKCTOCS Data Manager; Figure 4-4, Figure 4-5, A-1, A-2, Studies in Chapter 7, 8 and 9.
- The study described in Chapter 8 was designed by the UKCTOCS Data Manager and members of the Ultrasound Subcommittee
- Data related to Kappa calculation in Chapter 8 – Matthew Burnell, Women’s Cancer Statistician; Table 8-1

Co-authors of papers

The work described in the following chapters has been written-up as scientific papers intended for publication and a number of co-authors have been involved in this process as indicated:

- Chapter 7: Capability of the Image Workflow for TVS Quality Control
 - A. Franchini, C. Jones, A. Ryan, S. Campbell, A. Gentry-Maharaj, G. Fletcher, J. Kalsi, I. Jacobs, U. Menon.
- Chapter 8: Audit of TVS Examinations by a Team of Experts
 - A. Ryan, A. Gentry-Maharaj, G. Fletcher, M. Burnell, N. Asmo, M. Seif, A. Ferguson, G. Turner, C. Brunell, K. Ford, R. Ranger, S. Campbell, C. Jones, I. Jacobs, U. Menon.
- Chapter 9: Automated Collection of TVS Quality Control Metrics
 - S. Campbell, A. Franchini, O. Blyuss, A. Ryan, C. Jones, A. Gentry-Maharaj, G. Fletcher, J. Kalsi, S. Skates, M. Parmar, I. Jacobs, U. Menon

The co-authors collaborated in the design of the above studies and the generation of the data as well as its interpretation. All authors contributed to the editing of the papers parts of which are reproduced in the above chapters.

The Institute of Women's Health has a custom of giving authorship credit to anyone who has had any involvement in the generation of a paper submitted for publication. The level of individual contribution is not usually identified, though by convention in a multi-authored paper based on a student thesis or dissertation the student appears as the first author and their secondary supervisor as last author. Consequently in the cases of the papers written in respect of the work described in Chapters 7, 8, 9 there is no attempt to identify levels of individual contribution and papers are presented as a team effort. In order to provide consistency between claims of authorship the description of author contribution is similar in both the papers and the thesis. However, none of the papers had been accepted for publication by the thesis submission date (1/7/16), so sole ownership of the copyright at this date is claimed by William Stott as prescribed by the Copyright, Designs and Patents Act 1988 (excepting Tables and Figures, as acknowledged). The legal basis of this claim is that he is the original creator and first owner of the works and was responsible for creating all updates of the works including the final drafts prepared for publication. Whilst it is acknowledged that a number of other people made suggestions for improving the works, including in some cases suggestions for alternative wording, these did not constitute a substantial individual contribution to the final text of the papers as might warrant a claim for joint ownership of the copyright. Ownership of sole copyright is also claimed by William Stott in respect of Chapters 7, 8, 9 on similar grounds as they are derived from these papers and also benefit from review by academic supervisors without substantial contributions being made to the final text.

General Assistance from Supervisors

In addition to the above the thesis project and the thesis itself benefited from supervision provided by C. Jones and U. Menon. This included commenting on draft chapters and making suggestions for changes to the text.

Introductory Chapters

Chapter 1: Problem Domain of the Thesis

Introduction

This thesis is about the use of software tools to implement quality control of ultrasound images obtained from a large clinical trial – the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS). The primary outcome of this trial was ovarian cancer death by 31st December 2014. The trial was designed “*to establish the effect of early detection by screening on ovarian cancer mortality*”, as stated by Jacobs et al¹. Volunteers were recruited and randomly allocated to three groups; the ultrasound screening group, a multimodal screening group and a control group that received no screening. The images considered by the thesis were selected from those recorded during the scanning of women in the ultrasound screening group, as detailed in Appendix A. Such scanning is considered central to screening strategies for ovarian cancer², but is acknowledged to have a substantial subjective element³.

By way of an introduction to the research the problem is stated and its context given in terms of the impact of ovarian cancer, the anatomy of the uterine adnexa and the nature of the UKCTOCS with particular regard to the ultrasound screening part of the trial and the images it produced.

Problem Statement

Transvaginal sonography (TVS) is core to any ovarian cancer screening strategy yet independent and objective quality control (QC) metrics for this procedure are not routinely obtained. The research question this thesis aims to answer is: *Can software tools be developed to collect independent and objective QC metrics for TVS exams by routinely analysing the images generated for ovary measurement?* Such tools have application in studies of images from thousands of TVS exams performed by UKCTOCS and allow conclusions to be made about the quality of this scanning.

Importance of Quality Control in Ovarian Cancer Screening

The research has particular relevance in respect of UKCTOCS as it demonstrates the feasibility of automating the QC of TVS scanning to drive quality improvement which would address the key practical issue of monitoring and controlling scan quality in any future national ovarian cancer screening programme.

The research has translational potential for improving the quality of TVS in the National Health Service (NHS) which may assist in the early diagnosis of the disease and result in improved outcome for some women in respect of the >1.2 million scans performed annually⁴. It might also encourage further research to investigate whether the differences in ovarian cancer five-year survival rates between the UK and other European countries can be attributed to TVS quality, so addressing one of the key challenges of ovarian cancer research that of improving early diagnosis of the disease.

Impact of Ovarian Cancer

Ovarian cancer is the fourth most common cause of cancer death in UK women with 7,029 new cases and 4,271 women dying of the disease in 2012⁵. Ovarian cancer accounts for more deaths than all the other gynaecological cancers put together. It impacts older women in particular with incidence rates increasing significantly after the onset of the menopause and peaking at 71 per 100,000 UK women in the 80-84 year age group⁶. Attempts to reduce the mortality rate of ovarian cancer have been

less successful than in other types of cancer such as cervical and breast. This reflects the relative lack of understanding about its nature, the challenges presented by its prevention and cure as well as the difficulties of diagnosis.

The European average five-year survival rate for ovarian cancer is 37.6% with a rate of 41.4% in Austria and 31% in UK⁷. The reasons for this discrepancy between the UK and other European countries are unknown. However, it is widely recognised that early diagnosis improves the prognosis significantly. For example, women diagnosed with early stage ovarian cancer have a five year survival rate of 90% compared with late stage diagnosis which has a five year mortality rate of 5%⁸. Nevertheless, the benefits of early diagnosis in some women must be balanced against the harm caused by unnecessary treatment of women who are misdiagnosed with the disease; over-diagnosis.

The UKCTOCS randomised controlled trial was funded to provide a definitive answer to the question as to *whether a national screening programme for ovarian cancer would save lives*. Any subsequent plan to implement a national screening programme for ovarian cancer would need to take account of the feasibility of obtaining objective and independent quality control (QC) information for the TVS scans which are core to ovarian cancer screening. This thesis aims to provide information in this regard.

Anatomy and Pathology of the Uterine Adnexa

The gross anatomy of the adnexa region, as described by Walker et al⁹, is shown in Figure 1-1. The principal organs and associated tissues can be readily identified in textbook views, but it should be kept in mind that an ultrasound image shows them in-situ. Therefore, their position and shape may be distorted by the presence of body fat, muscle and other organs such as the bladder and lower intestinal tract.

Ultrasound is also subject to distortion as discussed in Chapter 2 so the content of its images does not always correspond to the form of organs in the body.

TVS Examinations are primarily concerned with identifying neoplasms in the location of the ovary and endometrium (lining of the uterus), but other structures

form landmarks which are helpful in terms of providing information about the orientation of the ultrasound image:

- Cervix – Entrance to the endometrial cavity
- Uterus – Reproductive sex organ
- Fallopian tube – Ovum conduit allowing sperm fertilization and attachment to uterus
- Broad Ligament – Attached to the uterus, it supports the ovary
- Internal Iliac artery (and vein) – Blood supply for the reproductive organs

The ovary of a woman before her menopause is typically 3 cm x 1.5 cm x 1.5 cm (volume 3.5cm³). It is shaped like an olive and covered by a single layer of ovarian surface epithelium (OSE) cells. There is an inner medulla layer and an outer cortex layer which are mainly composed of spindle-shaped stromal cells as well as collagen within a reticulum network. The cortex is also well supplied with blood vessels and interspersed with follicles at various stages of development, but lacks any definitive boundary with the medulla. In contrast the stroma is separated from the surface epithelial cells by a distinct basement membrane and tunica albuginea. This provides a barrier that prevents the diffusion of bioactive compounds between the two tissue types.

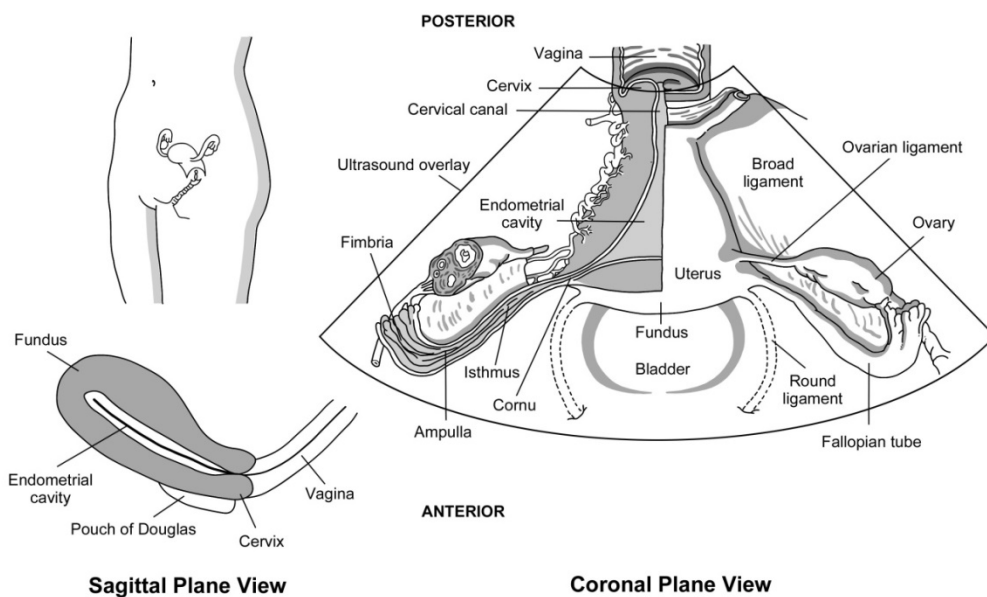


Figure 1-1: The uterine adnexa contains the ovaries, uterine tubes (also termed fallopian tubes or salpinges), the endometrial cavity and bladder as well as the small intestine, vessels, nerves and so forth.

The surgical removal of an ovary with the intention of preventing disease is termed a prophylactic oophorectomy and the removal of both ovaries and both fallopian tubes was termed a prophylactic bilateral salpingo-oophorectomy (PBSO), but is now commonly termed risk reducing salpingo-oophorectomy (RRSO). Such procedures may be carried out on women considered to have a high risk of ovarian cancer. This is particularly common in the case of women who have risk factors that include inherited mutations of the genes BRCA1 or BRCA2 which are associated with various forms of cancer including breast and ovarian, as reported by Randall and Pothuri¹⁰.

Primarily the ovary serves as an ovum-producing reproductive organ containing various follicles as well as an active Corpus Luteum. The ovum develops from a primary Oocyte contained in a Primordial Follicle. Post menarche a Primordial Follicle will grow first into a Primary Follicle and then into a Secondary Follicle before finally becoming a mature Graafian follicle which, during ovulation, ruptures to release the ovum so that it might migrate into the Fallopian (uterine) tube for subsequent fertilization. After releasing the ovum, the surface epithelium of the ovary repairs so encapsulating the follicle back into the ovary where it forms into a hormone producing Corpus Luteum which eventually degenerates into fibrous scar tissue called a Corpus Albicans (white bodies) as described by Tortora and Derrickson¹¹.

After the menopause the ovary stops releasing ovum and follicle development ceases which causes the ovary to shrink significantly as the stroma atrophies and collagen becomes more abundant. However, as mentioned by Kurman et al¹² androgens continue to be secreted by the stroma making the ovary a major source of testosterone and androstenedione in post-menopausal women which is why their removal is not without consequence. Such ovaries are harder to detect during TVS examinations and contain few discernible features except for abnormalities like cysts which are found in certain women.

Cancer in the Uterine Adnexa

A tumour (lesion) results from dysplasia which is an abnormal proliferation of cells. Typically, it is differentiated from surrounding tissue and forms a solid lump termed a neoplasm (new growth). Malignant tumours develop into forms of cancer whereas benign tumours are not associated with such diseases. The nature of an ovarian neoplasm may be predicted during clinical examination, but a definitive diagnosis requires histopathological analysis of actual tissue obtained by biopsy or surgery.

Ultrasound examination can reveal the presence of complex ovarian morphology which may be indicative of ovarian cancer. Timmerman et al¹³ identify a cyst with a septum, multiple cysts, solid areas inside the ovary, papilla or irregularities in the outer wall of the ovary in this regard. However, little is known about the features found in TVS images which indicate the onset ovarian cancer. For example, at present a clinician cannot predict whether a particular small cyst will develop into some form of complex morphology or will remain benign for the rest of the woman's life. The risk associated with the sort of interventions necessary to eradicate such a cyst may well be greater than the risk of it developing into ovarian cancer, particularly in the case of an older woman who is more likely to suffer from complications arising from treatment.

Diagnostic Accuracy of Tests for Ovarian Cancer

It is clearly important that any test for ovarian cancer produces reliable results in order to avoid unnecessary interventions arising from results that incorrectly predict the presence of disease (false positives) and missed opportunities for timely interventions arising from results that incorrectly fail to detect the disease (false negatives). Studies into the diagnostics accuracy of such tests are concerned about agreement between the index test and a reference test in respect of identifying a particular condition – i.e. ovarian cancer. Such studies generally measure the outcome of the new index against a reference test when performed sequentially on a series of patients selected at random using the same criteria. This allows the construction of a 2 X 2 table with the results of both tests on adjacent sides; see Table 1-1.

	Reference Test Positive	Reference Test Negative	Total
Index Test Positive	True Positive (TP) count	False Positive (FP) count	TP + FP
Index Test Negative	False Negative (FN) count	True Negative (TN) count	TN + FN
Total	TP + FN	TN + FP	Total tests performed TP + FP + TN + FN

Table 1-1: Test outcome results presented in terms of a 2 X 2 Table. Analysis of the values in such a table allows the production of measures of the test's performance such as accuracy, sensitivity and specificity.

Various measures of the index test's performance can be obtained from the analysis of the 2 X 2 table values:

- **Accuracy** is given as the sum of true positives (TP) and true negatives (TN) divided by the total tests performed.
- **Sensitivity** is given as the number of true positives (TP) divided by the sum of true positives (TP) and false negatives (FN).
- **Specificity** is given as the number of true negatives (TN) divided by the sum of true negatives (TN) and false positives (FP) .
- **Positive predictive value (PPV)** is the number of true positives (TP) divided by the total number of tests with positive results: $PPV = TP / (TP + FP)$.

Each of these measures provides a way to compare test performance, though there is evidence that likelihood ratios presented in non-technical language are more easily understood by clinicians allowing a better interpretation of the test. Florkowski¹⁴ reported that in respect of the two comments about Transvaginal ultrasound showing a pathological result compatible with cancer:

- *“A positive result is obtained twice as frequently in women with an endometrial cancer than in women without this disease”*
- *“The sensitivity of this test is 80%, its specificity is 60%”.*

The first comment was better understood by a sample of General Practitioners than the second even though there is a long tradition of reporting the diagnostic performance of clinical tests in terms of sensitivity and specificity.

Most diagnostic tests involve parameters that influence their sensitivity and specificity such that increases in sensitivity result in decreases in specificity and vice versa. For example, decreasing the level of CA-125 (or change in level) that determines the difference between a positive and negative test will result in a higher sensitivity, but at the cost of reduced specificity. Similarly, ultrasound examinations can be made more specific but less sensitive by making the criteria for abnormality more stringent. A Receiver-Operating Characteristic Curve (ROC) is often used to assist in the selection of a suitable cut-off point for such test parameters; see Figure 1-2.

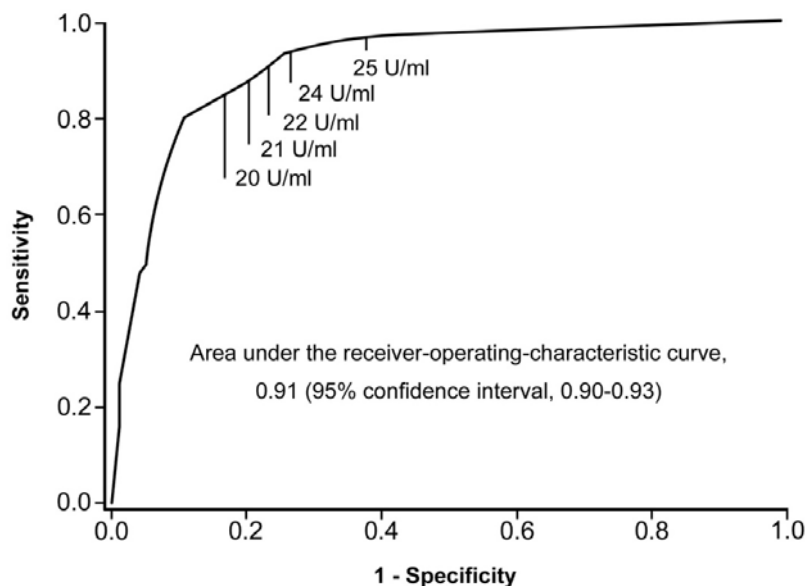


Figure 1-2: ROC for optimisation of test sensitivity and specificity. It can be seen that the selection of different cut-off values for CA-125 changes its sensitivity and specificity such that higher value allow the test to be more sensitive, but at the cost of it becoming less specific.

The area under the ROC curve is a measure of the test accuracy such that an excellent test provides very high sensitivity and is highly specific to give near perfect discrimination between people with and without the disease without regard to cut-off points for parameter selection. In this case the area under the curve would tend to 1, much like the curve in Figure 1-2. Conversely a worthless test would not allow any discrimination between people with and without the disease regardless of cut-off points for parameter selection, giving curve tending to a diagonal line with an area under it tending to 0.5.

Clinical Prediction of Ovarian Malignancy

Geomini et al¹⁵ performed a systemic review of the literature in February 2009 and concluded that the Risk of Malignancy Index (RMI 1) developed by Jacobs et al¹⁶ is the test of choice in the preoperative assessment of the adnexal mass. RMI 1 determines the risk of malignancy as an index value calculated from the product of serum cancer antigen 125 (CA125) level given in units per millilitre, the ultrasound scan score (0, 1 or 3) and the menopausal status (1 for premenopausal and 3 for postmenopausal). The ultrasound scan score reflects the number of points accumulated for certain features detected during examination; multilocular cysts, solid areas, metastases (further malignant tumours resulting from the spread of cancerous cells from the original tumour site), ascites and bilateral lesions. The International Ovarian Tumour Analysis (IOTA)¹³ group has helped to ensure consistency in such scoring by standardizing the terms, definitions and measurement of sonographic features for adnexal tumours, particularly in terms of:

- **Lesion** – part of an ovary or an adnexal structure that is judged to be inconsistent with normal physiologic function. Lesions in both ovaries are termed bilateral.
- **Septum** – a thin strand of tissue running across the cyst cavity from one internal surface to the counter-lateral side.
- **Solid tissue** – tissue exhibiting high echogenicity not associated with wall thickening, normal ovarian stroma or regular septa.
- **Solid papillary projections** –solid projections into the cyst cavity from the cyst wall with a height greater than or equal to 3mm.
- **Complex Cysts** – irregular inner wall in respect of the inner wall of a simple cyst or the outer wall of a solid component or having content that can be described as appearing like ground glass (homogeneously dispersed echogenic cystic contents) or haemorrhagic (internal thread-like structures; star-shaped, cobweb-like, jelly-like).
- **Ascites** – accumulation of fluid in the peritoneal cavity outside the Pouch of Douglas (POD). This indicates normal absorption has been disturbed, possibly due to the presence of free-floating cancer cells.

In addition when colour Doppler imaging has been performed it can be described in terms of a subjective semi-quantitative score of blood flow; 1 - no blood flow can be found in the lesion; 2 - detection of only minimal flow; 3 - presence of moderate flow; 4 - adnexal mass appears highly vascular with marked blood flow. Using the above terms Timmerman et al¹³ qualitatively classify the lesions detected during a TVS examination as:

- **Unilocular cyst** – without septa and without solid parts or papillary structures
- **Unilocular-solid cyst** – unilocular cyst with a measurable solid component or at least one papillary structure
- **Multilocular cyst** – a cyst with at least one septum but no measurable solid components or papillary projections
- **Multilocular-solid cyst** – multilocular cyst with a measurable solid component or at least one papillary projection
- **Solid tumour** – where the solid components comprise 80% or more of the tumour when assessed in a two-dimensional section. A solid tumour may contain papillary projections protruding into the small cysts of the solid tumour.
- **Not classifiable** – because of poor visualization.

These lesion classifications were initially developed to standardise the categorical variables used in logistic regression models and machine learning systems for predicting ovarian malignancy as reported by Timmerman et al¹⁷. However, they are also useful for standardizing the subjective evaluations made by sonographers of grey-scale ultrasound images, a technique commonly referred to in the literature as pattern recognition. This can lead to an almost conclusive diagnosis of certain types of adnexal pathologies and the recognition of many other types as reported by Sokalska et al¹⁸. Other researchers have published similar work in an attempt to standardise the diagnosis of benign and malignant adnexal masses during TVS examinations. For example Brown, Dudiak and Laing¹⁹ produce the following useful descriptions of common benign adnexal masses:

- **Simple cyst:** lack any solid area or septa and characterised by a thin smooth wall containing fluid (anechoic) with distal acoustic enhancement^a. Subtypes include: inclusion cyst, parasalpingeal cyst and normal ovum.
- **Endometrioma:** typically appear as complex unilocular or multilocular cysts, with a ground glass appearance. They can partially or almost completely replace the normal tissue, though seldom exceed 15cm in diameter.
- **Dermoid (mature cystic teratoma):** usually have a hyperechoic area with distal acoustic shadowing.
- **Fibroma:** germ cell stromal neoplasm which appears as a solid homogeneous or heterogeneous mass. They often exhibit marked acoustic shadowing
- **Paraovarian cyst:** similar to a simple cyst, but located outside the ovary usually in the broad ligament.
- **Hydosalpinx:** an extraovarian mass which takes the form of a tubular shaped cystic structure with indentations on opposite sides of the wall – a '*waist*'.
- **Peritoneal inclusion cyst:** cystic masses external to the ovary typically with septa and exhibiting blood flow when colour doppler is applied.

It is important to discriminate between malignant and benign adnexal masses at the clinical prediction stage because the rupturing of a Stage 1 ovarian cancer during a subsequent investigation or treatment may worsen the prognosis for the patient, as reported by Vergote et al²⁰. It is for this reason that accurate diagnosis by TVS is important.

Pathological Assessment of Ovarian Malignancy

Normal cells become cancerous as a result of errors in the replication of DNA during the mitotic phase of cell division. This causes them to lack differentiation, suffer uncontrolled division and growth as well as a failure to respond to the normal mechanisms of programmed cell death – i.e. apoptosis or autophagy. Some

^a Distal acoustic enhancement is a sonographic artefact that arises when sound waves pass through a significant area of homogeneous (anechoic) fluid. There is less attenuation due to the absence of reflections interfering with the progression of the sound waves from the transducer and accordingly reflections from tissue distal to the fluid filled area appear more pronounced (brighter) than would be the case if the sound travelled through material that caused a degree of reflection before meeting this same tissue.

neoplasms remain benign so cannot be classified as cancerous, but others become malignant and cause disease.

Pathological diagnosis provides a primary classification of adnexal masses (tumours) as either malignant or benign as a result of histochemical or immunophenotype assay. Histochemical assay uses special stains and electron microscopy to identify cells in tissue, but increasingly immunophenotyping is used which involves the visualization by flow cytometry of antigens directed against given proteins known to be differentially expressed by malignant cells. These advances have resulted in significant changes in the understanding of the tumour biology of ovarian cancer over recent years leading to revisions of the way such tumours are classified both by type and stage of development.

Classification of Ovarian Tumour Types

Previous classification of ovarian tumour types assumed the point of origin for the malignant ovarian epithelial tumours (carcinomas) that account for approximately 90% of incidences of ovarian cancer was the mesothelial surface of the ovary. However, it is now recognised that the point of origin for many such tumours cannot be identified with any certainty. This new understanding is reflected in the World Health Organisation (WHO) revisions to the classification of ovarian tumour types as summarised by Meinhold-Heelein et al²¹. The basic classification of tumour type remains, but has the addition of a further Seromucinous type giving six major types which can be described in terms of their morphological, histopathological or immunohistochemistry as:

- **Serous** – considered as serous borderline ovarian tumours (SBOT) when more than 10% of the tumour cells are serous cancer type. They are considered malignant if they have intruded and damaged adjacent tissue or have spread elsewhere. Serous tumors are sub-classified as low grade serous carcinomas (LGSC) or high grade serous carcinomas (HGSC) depending upon the respective absence or presence of the p53 protein as identified by immunostaining.
- **Mucinous** – considered as mucinous borderline ovarian tumours (MBOT) when more than 10% of the tumour cells are mucinous cancer type and are

considered malignant if they have intruded and damaged adjacent tissue or spread to other organs.

- **Seromucinous** – considered as serous mucinous borderline ovarian tumours (SMBOT) when more than 10% of the tumour cells are serous cancer type (like SBOT) and a third are associated with endometriosis by virtue of ARID1A mutations. They are considered malignant if they have intruded and damaged adjacent tissue or spread to other organs.
- **Endometrioid** - considered as endometrioid borderline ovarian tumours (EBOT) when more than 10% of the tumour cells are endometrioid cancer type as identified by ARID1A and PIK3CA mutations as well as loss of PTEN heterozygosity. They are considered malignant if they are morphologically heterogeneous and have intruded and damaged adjacent tissue or spread to other organs. Further subtypes of EBOTs are identified.
- **Clear cell** – considered as clear cell borderline ovarian tumours (CCBOT) similar to clear cell adenomas. Such tumours are considered malignant if they have papillary structures, solid components with fibrous/connective tissue in the stroma or ground glass appearance.
- **Brenner** – composed of cell nests of various sizes characterised by transitional cell differentiation. Brenner borderline ovarian tumours (BBOT) have an extensive proliferation of epithelial cells and are much larger than benign Brenner tumors. They are considered malignant if there is destructive invasion of the stroma.

Classification of Ovarian Tumour Stage

The stage classification of ovarian cancer has been subject to recent revision to reflect changes in understanding about its development as reported by Prat²² on behalf of the International Federation of Gynecology and Obstetrics (FIGO) committee. Staging is the process of defining the development of cancer by the extent to which it is deemed to have advanced from its initial form. However, this task is complicated by tendency of certain types of ovarian cancer, like HGSC, to evolve in such a way that they have concealed the primary site by the time they are detected.

The following stages are now defined by FIGO:

- **Stage I:** Tumour confined to the ovaries or fallopian tube(s). Further sub-stages are defined to differentiate between tumours in one (1A) or both (1B) ovaries / fallopian tubes and tumours that have ruptured either before (IC2) or after (IC1) surgery or tumours located on ovary or fallopian tube surfaces (IC2) and the presence of malignant cells in the ascites or peritoneal washings (IC3).
- **Stage II:** Tumour in one or both ovaries / fallopian tubes which have spread to the pelvic region below the pelvic brim. Alternatively the tumour is located in the membrane that forms the lining of the abdominal cavity – i.e. primary peritoneal cancer.
- **Stage III:** Tumour in one or both ovaries / fallopian tubes (or primary peritoneal cancer) which cytological or histological analysis shows has spread to a) the peritoneum above the pelvic brim (sub-stage IIIA2) or b) retroperitoneal lymph nodes (sub-stage IIIA1). Further sub-stages are defined for peritoneal metastasis beyond the pelvic brim that is less than 2cm (IIIB) or more than 2cm (IIIC). Retroperitoneal lymph nodes metastasis may be present in any of the above sub-stages.
- **Stage IV:** Tumour in one or both ovaries / fallopian tubes (or primary peritoneal cancer) with distant metastasis excluding those found in the lining of the abdominal cavity. This includes patients with parenchymal liver or spleen metastases and those with metastases outside the abdominal cavity.

Although the term ‘stage’ implies a progression of the disease from Stage I to Stage IV, it is accepted that most ovarian cancers are HGSCs that typically present at Stage III with 84% of those presenting at IIIC. Patients staged as IIIA1 have no peritoneal involvement and can be further sub-staged according to whether or not the tumour is less than or equal to 10mm; respectively given as IIIA1(i) and IIIA1(ii).

The correct classification of ovarian tumours is necessary to decide appropriate treatment of the patient. It is also helpful in terms of screening as certain types are more associated with late detection than others. In this respect ovarian tumours of the high grade serous histology type are particularly significant as they are rarely detected in their early stages and are a major cause of mortality²³.

Development of Ovarian Malignancy

Increasingly it is clear that there is a dualistic model for origins of ovarian cancer, as reported by Kuman and Shih²⁴. They cite evidence from both morphological and molecular studies to suggest that there are two forms of ovarian cancer which can be considered almost as separate diseases. The Type II form is characterised by high grade serous carcinoma (HGSC) tumours arising in the main from intraepithelial carcinomas in the fallopian tube and can be further subdivided into various morphologic and molecular subtypes. Type II ovarian cancer is considered to develop rapidly and is highly aggressive with a poor prognosis such that it accounts for 90% of all ovarian cancer deaths. . The Type I form is characterised by other invasive epithelial cancer histotypes (endometrioid, clear cell, mucinous, low grade serous) developing from benign extraovarian lesions that implant on the ovary and subsequently undergo malignant transformation. Type I ovarian cancer is considered to develop slowly and is relatively indolent with an excellent prognosis if detected at early stage, though advanced stage tumours have a poor outcome.

Brown and Palmer²⁵ have attempted to model the natural history of high grade serous ovarian cancer (Type II as defined by Kuman and Shih) on the basis of a comprehensive analysis of published data on serous cancers discovered at risk reducing bilateral salpingo-oophorectomy in BRCA1 gene mutation carriers. They assumed exponential growth and produced different growth rates for early (stage I and II) and advanced (stage III and IV) progression of the disease.

The Brown and Palmer model suggests the potential of annual screening for ovarian cancer by proposing the genesis of ovarian cancer may occur as many as eight years before diagnosis and that the window of opportunity for taking action to save life after the detection of a carcinoma in situ stage I or II is 4.3 years (95% CI 2.6 to 6.9 years). However, it was anticipated that for such screening to achieve even 50% sensitivity it would need to detect tumours of 1.3cm or less in diameter– i.e. tumours only slightly larger than the 1.0cm diameter they considered as the lower limit for visualization during gross examination (TVS). The model also highlights the explosive growth of advanced stage serous ovarian tumours which are estimated to double in volume every 2.5 months, compared with a doubling every 4 months for early stage tumours. This suggests a significant difference in diameter between

tumours that would need to be found in screening (range 1.0–1.3cm) and those typically detected during diagnosis (median 3.0cm) which are anticipated to have reached stage III or IV a year before presentation.

To illustrate the different level of challenge between a TVS examination given as a screen for high grade serous ovarian cancer compared with when it is used as for diagnosis, consider the cases of two women with the disease; one having an annual screen and the other having a sporadic diagnosis as consequence of presenting with symptoms of the disease. In the case of the woman being screened, the examination would be need to be performed with considerable skill to allow detection of a very small abnormality in the ovary consistent with an early stage tumour that was too small to be found in the previous year's screen. However, in the case of the woman undergoing a sporadic diagnosis, the examination would not require the same level of expertise as the abnormality would be much easier to identify. This is because the onset of symptoms is typically associated with later stage development of the disease when the tumour is large and doubling in size every few months. Indeed, the Brown and Palmer model predicts that it is likely to have been detectable by TVS ($>1.0\text{cm}$ in diameter) for several years before diagnosis.

Brown and Palmer found that a screening test with 80% detection sensitivity was dependent on identifying tumours of less than 0.4cm in diameter and that that annual screening might achieve a 50% reduction in serous ovarian cancer mortality if tumours of 0.5cm diameter were detectable. They conclude that although a window of opportunity for the early detection of such cancers lasts for several years, it will be challenging to develop an annual screening test with the levels of sensitivity and specificity needed to take advantage of it. They suggest that an effective screening programme would require the detection of tumours much smaller than is currently achieved using biomarkers like CA-125, so call for the development of alternative approaches to testing.

Brown and Palmer do not make any claims about their model in respect of ovarian cancer types other than those of serous histology. They also emphasize that their data is based only on women with the BRCA1 genetic mutation, but consider that their model is reliable in the case of women without such mutation. The conclusions of Brown and Palmer have not been replicated by other similar studies and their paper

predates even the initial work by Kurman and Shih²⁶ on the dualistic model of ovarian cancer, so no account is made of the disease's possible classification into Type I and II or the findings from more recent morphological and molecular studies.

Although Kurman and Shih do not cite Brown and Palmer in their 2016 paper, both papers express doubts about the effectiveness of TVS and CA-125 testing in ovarian cancer screening and call for the development of alternative biomarkers. Kurman and Shih observe that the goal of ovarian cancer screening trials to date has been the detection of Stage I disease in the ovary as this is associated with 90% survival compared with 30% survival for advanced stage disease (Kobel et al²⁷ are cited). They also note that such trials have failed to demonstrate a significant survival benefit; references are made to papers reporting mortality benefit from the Prostate, Lung, Colorectal and Ovarian (PLCO) trial²⁸ as well as UKCTOCS¹. By way of a possible explanation they suggest that stage I neoplasms are associated only with Type I disease and present as large cystic masses that can be easily detected by TVS screening. However, the HGSC tumours which characterise the Type II form of the disease and account for 90% of deaths occur at inception in the fallopian tubes when the neoplasm is already classified as stage II and is unlikely to be detectable using either TVS or CA-125. Therefore Kurman and Shih conclude that any major impact on mortality will only arise from the development of more sensitive screening tests able to detect a HGSC tumour early in its evolution when it is confined to the fallopian tubes. In this respect their findings do not support the conclusions of Brown and Palmer which suggest that screening tests for HGSC tumours in the ovary of size 1.0cm to 1.3cm diameter might achieve 50% sensitivity.

The preliminary UKCTOCS findings as reported by Jacobs et al¹ show a mortality benefit arising from screening, but it is small and found mainly in the multimodal screening arm (MMS). This suggests annual screening may be worthwhile as the development of ovarian cancers present a window of opportunity for their early detection that lasts for several years, as proposed by Brown and Palmer. The lack of mortality benefit found in the ultrasound arm (USS) could be explained by: a) the detection of only the Type I form of the disease accounting for just 10% of deaths as proposed by Kurman and Shih, b) the higher level of performance needed when TVS is given as an annual screen for ovarian cancer rather than as a diagnostic. In this

latter regard the slow progression of a small benign neoplasm into small Type I tumour which can be detected before it develops into late stage disease is associated with excellent prognosis, so yielding an effective mortality benefit. However, as such ovarian cancer screening has only been performed on a limited basis in a few clinical trials there is a lack of information about how the necessary performance improvements in TVS scanning might be realised in large-scale use.

The quantification of any gain in TVS performance depends on good quality control and for this reason tools like those developed for this thesis have an important role in obtaining high quality data from clinical trials and other sources. This will allow research into the use of ovarian cancer screening with the aim of detecting ovarian malignancy at the earliest possible stage of its development when treatment is most likely to prove effective, though according to Kurman and Shih any impact on mortality is likely to remain limited.

UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS)

The UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS)²⁹ is a multicentre randomised control trial involving postmenopausal women from 13 trial centres (TC) in England, Wales and Northern Ireland. Its main aim is the assessment of screening on mortality from malignant neoplasms of the ovary as defined by code C56 of the International Classification of Diseases (10th Revision). The trial started in 2001 and it is anticipated that the final results of UKCTOCS will be reported in 2018, though the initial results of its primary end point were reported on 17th December 2015 as previously mentioned.

Recruitment and Randomisation

Trial volunteers were recruited between 2001 and 2005. After confirmation of eligibility 202,638 volunteers were enrolled and randomly allocated in a 2:1:1 ratio respectively to:

- Control group: received no screening (n=101,359)

- Ultrasound Screening (USS) group: received annual screening using transvaginal scanning (TVS), though a transabdominal (TAS) scan was carried out when TVS was not acceptable to the volunteer (n=50,639)
- Multimodal Screening (MMS) group: received annual screening using serum CA125 interpreted by the Risk of Cancer (ROC) algorithm with an ultrasound scanning given as a second-line test (n=50,640)

Further details about the recruitment and allocation of the volunteers to the above groups are given in Appendix A.

Inclusion and Exclusion Criteria

The trial protocol (see Appendix B) specified the inclusion criteria as women aged at recruitment between 50-74 years and postmenopausal. Women with previous ovarian malignancy, bilateral oophorectomy, active non-ovarian malignancy, increased risk of familial ovarian cancer or participants in other ovarian cancer screening trials were excluded.

Ultrasound Screening

Ultrasound screening started in July 2001 and ended in December 2011. During this period 48,250 volunteers in the USS group received one or more scans. A total of 300,027 annual TVS scans were given to these volunteers. This number excludes annual scans given using a combination of TVS and TAS, or TAS only scans, or scans given as second-line test to volunteers in the MMS group. It also excludes repeat scans following either an unsatisfactory annual scan (termed level 1 repeat) or an annual scan that detected an abnormality (termed level 2 repeat); see Classification of Examinations.

TVS Examination

The TVS examination is performed by a sonographer, though this term encompasses people with a range of professional qualifications from mid-wife to consultant gynaecologist. Details of the equipment used for TVS examination are given in Chapter 2 and its quality assurance is discussed in Chapter 3. The procedure adopted by UKCTOCS for ovarian cancer screening was modelled on the procedure

generally used in the NHS for ovarian cancer diagnosis as more appropriate models for such screening have yet to be developed. In this respect UKCTOCS has some potential to influence future work by disseminating information about both its successes and failures. The extracts of the trial protocol relevant to ultrasound screening are given in Appendix B.

The average time allowed for a TVS scan is 15 minutes and there is pressure for sonographers to complete their list of volunteers within a given clinical session. At the start of an examination the sonographer inserts an ultrasound probe into the woman's vagina and then guides its movement using a sense of touch combined with the real-time video image displayed on the machine's monitor so as to view features of interest. The image is displayed in two dimensions corresponding to a cross-section of the adnexa displayed in an arc from the tip of the probe to the extent of the ultrasound echoes in the direction of the probe. Type of image produced is shown in Figure 1-3, though this is just a snap-shot of the real-time video image presented to the sonographer on the ultrasound machine display. In this case a cross-section of the uterus is shown in the middle (left) and the probe tip is at the top of the image.

As the scan proceeds, the sonographer is able to build-up a mental map of the structure of the adnexa by manipulating the probe, in much the same way that you might explore a large dark cave by flashing a torch light with a narrow beam around the rock surfaces. This allows the sonographer to home-in on features of interest and apply techniques like external palpation of the abdomen in order to bring organs obscured by bowel or other structures into view. Experienced sonographers have considerable skill in manipulating the probe in order to build-up these mental maps. This is prerequisite to obtaining a good view of the ovaries which may be difficult to find, particularly due to their small size in postmenopausal women. In addition to confirming the ovaries and endometrium have normal morphology the sonographer also checks for abnormalities like the presence of Ascites or fluid in the Pouch of Douglas. The skill of the sonographer in relaxing the volunteers during an unpleasant and invasive procedure is an important factor in achieving a satisfactory examination, but other factors influence scanning such as the recent fluid intake by the volunteer as well as her age, history of adnexal surgery (hysterectomy,

oophorectomy, tubal ligation), age at menopause and body mass index (BMI) as reported by Sharma et al³⁰.



Figure 1-3: Static image of a section of the uterus. The sonographer has measured endometrial thickness as 0.63cm by placing the callipers between the boundaries of the endometrium.

The UKCTOCS trial protocol requires the sonographer to record the thickness of the endometrium as well as the size of each ovary in terms of its dimensions in two orthogonal planes. This requires the sonographer to take a screenshot of the video image and place calliper marks on the feature being measured so that the machine can then calculate the distance and display its value at the bottom of the image; see Figure 1-4. These screenshots are termed ‘static images’ and can be stored as computer files as described below. The protocol specifies that static images should be recorded showing the transverse section (TS) and longitudinal section (LS) of the left and right ovary as well as a cross section of the endometrium, though other image types are expected for certain conditions that might be encountered during an exam.

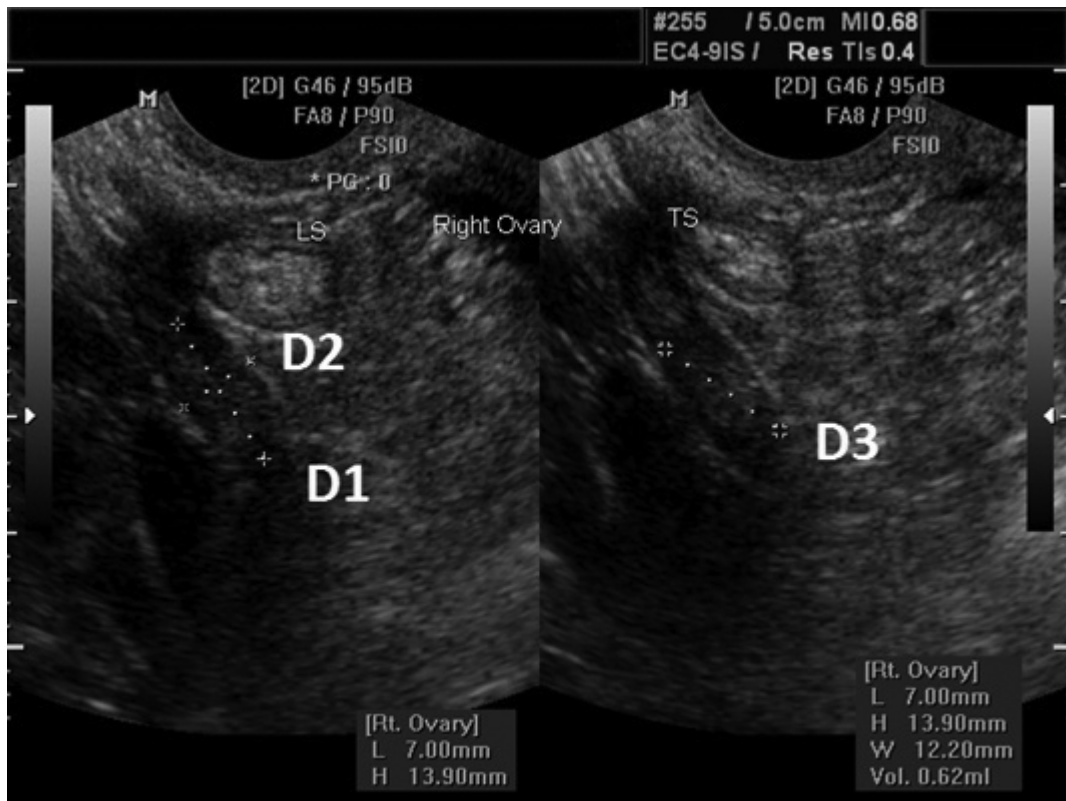


Figure 1-4: Measurement of ovary dimensions. The sonographer has placed callipers on the longitudinal (left) and transverse (right) sections of the ovary to measure its dimensions in three axes. The annotations added to the image reflect their standard naming; D1, D2, D3.

Images Used for Studies

Static images from TVS exams performed by UKCTOCS are 640 pixels wide and 480 pixel high with 8 bits of grey-scale information per pixel. They are stored on the Trial Centre (TC) ultrasound machines as computer files in Digital Imaging and Communications in Medicine (DICOM) format with 24 bits per pixel to allow the images to contain full colour elements, such as Doppler flow as described in Chapter 2. These DICOM files are similar to standard bitmap files, but have a more complex header which allows storage of information such as the volunteer's name, UKCTOCS identifiers, etc. A more complete description of this format is given in Appendix H.

The static images are transferred from the TCs to the trial Coordinating Centre (CC) where they are processed and stored in the Ultrasound Record Archive, a computer system developed specifically for UKCTOCS; see Chapter 4. The process used to transfer these DICOM files is detailed in the workflow described in a subsequent

section of this chapter. The URA contains static images from 216,152 of the 300,027 (72%) annual TVS examinations performed by UKCTOCS. Appendix A contains a diagram showing how these images were subsequently selected for the studies described in Chapters 7, 8 and 9.

Completion of the Scan Report Form

The UKCTOCS trial protocol prescribes the form of the scan report (Appendix C). It is completed by the sonographer during the scan or immediately afterwards and features a large number of checkboxes for binary data, spaces for the entry of figures and text as well as some spaces for a few lines of free-form text. In addition to information about visualization, ovary dimensions and morphology the scan report also records the endometrial thickness, data about cysts, the results of any Doppler study as well as other information having clinical or research value.

A key part of the scan report is concerned with documenting the type of ovary visualization achieved as this information is important for the purposes of determining the quality of the ovarian cancer screening. The visualization types defined by the UKCTOCS protocol are given in Appendix B in terms of:

- Ovary seen.
- Ovary not seen due to previous oophorectomy.
- Ovary not seen, but iliac vessels seen (good view).
- Ovary not seen and iliac vessels not seen (poor view).

Ovary visualization is a key consideration when screening for ovarian cancer due to the need to detect relatively small morphological changes in the ovary, unlike the gross changes that are likely to be found when TVS is used as a diagnostic for ovarian cancer (see previous discussion). Therefore it cannot be definitively claimed that a woman doesn't have any indication of ovarian cancer unless both her ovaries are seen and examined by the sonographer (or the remaining ovary in the case of a woman with an oophorectomy). However, it is commonly accepted that the visualization of ovaries in some women is difficult. Accordingly, UKCTOCS permits examinations to be deemed satisfactory if an ovary is not visualised, but more than 4cm of iliac vessel is seen. This is considered a 'good view' and indicates that the sonographer was looking in the right place, but the ovaries were not imaged

by ultrasound for some reason; see Chapter 2 for possible explanations. The rationale for deciding that such ovaries do not contain small tumours of the size that need to be detected for effective screening seems to depend on an assumption that the woman's ovary has no such tumour if the ovary itself is too small to be seen – i.e. <1.0 cm as suggested by Brown and Palmer²⁵.

As previously mentioned, the trial protocol obliges the sonographer to record the size of all visualised ovaries two orthogonal planes. The scan report contains three fields for this information corresponding to the longest line marking the extent of the boundary in the longitudinal section, the maximum extent of the boundary at 90 degrees to this line, and longest line marking the extent of the boundary in the transverse section. These measurements are called D1, D2, and D3 respectively; see Figure 1-4. However, they were not explicitly identified as such in the scan report which resulted in the Trial Management System (TMS) containing values for these fields in a variety of orders; see Chapter 7.

The presence in the TMS of values for D1, D2 and D3 in respect of an ovary is important because it provides evidence that the sonographer not only visualised the ovary during the given TVS exam, but also recorded the static images (LS and TS) required for its measurement. Therefore these images must contain calliper marks identifying the feature the sonographer considered to be an ovary. This allows a subsequent review of such images by experts in gynaecological scanning in order to obtain independent agreement that the features within the calliper marks are indeed ovaries.

In addition to recording ovary dimensions and endometrium thickness, the scan report also requires the sonographer to describe the ovary morphology in terms of its being observed as being normal, having a simple cyst or having complex morphology. This information is important as it forms part of the information used to classify the TVS examination.

Classification of TVS Examinations

The scan report provides the information necessary for an algorithm implemented by the TMS to classify the scan, as described in Appendix D, in terms of it being:

- **Abnormal:** any of the following conditions found:
 - Fluid in the Pouch of Douglas or Ascites detected in adnexa
 - Any ovary containing either complex morphology, or more than one simple cyst, or a simple cyst of volume greater than 60 cm³
- **Unsatisfactory:** not abnormal and both of the following conditions satisfied:
 - One or more of ovaries present were not seen
 - Iliac vessels not seen in place of an ovary not seen (one or both)
- **Normal:** a scan that was not abnormal or unsatisfactory. This means the ovaries were considered to have normal morphology and no other abnormalities in the adnexa were detected.

A classification of unsatisfactory triggers a level 1 repeat scan whereas a classification of abnormal triggers a level 2 repeat scan. The procedure for repeat scanning is defined by the trial protocol; see Figure B-1 in Appendix B. However, a scan classified as normal is not repeated as there is an assumption that the volunteer's ovaries did not contain a malignant tumour that could be detected by ultrasound. This means the volunteer will receive her next scan in the following year as a routine annual examination.

Scanning Workflow

The management of the UKCTOCS trial was facilitated the TMS. In respect of the ultrasound arm it was responsible for collecting and storing the scan reports as well as performing administrative tasks such as issuing appointment letters to volunteers, producing reports, etc. It was also used to collect and store trial data common to both arms of the trial such as the full clinical history of each volunteer (maintained until death).

The workflow for completing a TVS scan is summarised in Figure 1-5. In terms of the administration of a clinic for ovarian cancer screening, it starts with the TMS generating an appointment letter asking the volunteer to attend an examination at their local trial centre. When the volunteer arrives at the clinic a scan report form is

printed and the volunteer's details are handwritten into the appropriate sections. The form is then given to the sonographer so some of the details can be typed into the ultrasound machine for displaying in the images and storing as metadata in the DICOM headers of recorded files. After the start of the examination, observations are recorded on the form by the sonographer. This involves ticking check boxes, entering figures and text, or making free-text comments. After the scan, the information in the form is transcribed into the TMS using a form presented by a web browser, usually on a PC located in the TC office. For the majority TVS examinations this was job is done immediately after the completion of the scan or after the end of the clinic.

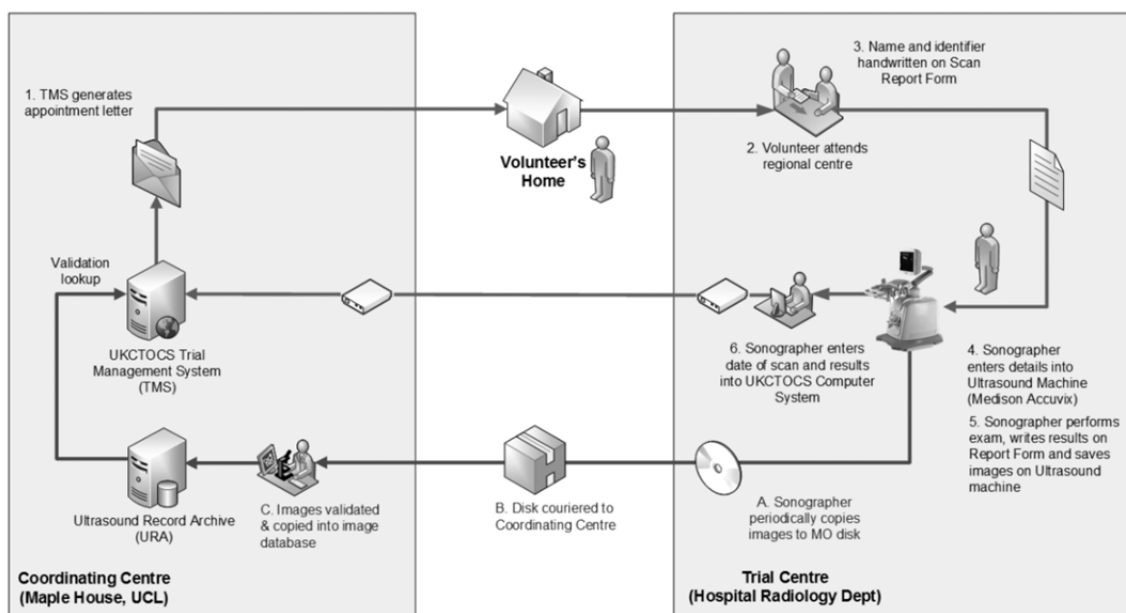


Figure 1-5: UKCTOCS Workflow for performing a TVS Scan. A letter is sent to the volunteer's home giving details of their next appointment for screening. A TVS scan is performed when the volunteer attends this appointment and details from the scan report form are entered into the TMS. In addition the static images used to measure the ovaries and endometrium are stored and periodically sent to UKCTOCS coordinating centre for archiving in the URA.

In addition to describing the administration of screening clinics, the scanning workflow also shows how the examination images from the 13 trial centres were periodically transferred to the coordinating centre in London. This process required the trial centre sonographers to copy onto a magneto-optical (MO) disk all the DICOM files that had been created on the ultrasound machines during prior TVS examinations; a task usually done weekly. The MO disk would then be sent to the

coordinating centre by courier where it enters a queue pending entry into the Ultrasound Record Archive (URA). The processing and archiving performed by the URA is described in Chapter 4.

Statutory and Trial Regulations for UKCTOCS

The operation of the trial was subject to UK law as well as academic, clinical and research regulations. The law and regulations that had particular relevance to the ultrasound arm of the trial were the Data Protection Act (DPA) and Ethics approval.

Data Protection Act Compliance

If the images are pseudo-anonymised before storage as described in Chapter 4, then it is understood that they are no longer categorised as personal medical data so are not subject to the DPA. There are no specific legal requirements for the storage of images from UKCTOCS, though there is a general requirement to take reasonable precautions to ensure they are not lost and cannot be accessed without appropriate authorisation.

Ethics approval

The UKCTOCS study was approved by North West Multicentre Research Ethics Committee 21/6/2000; MREC reference 00/8/34. It is registered as an International Standard Randomised Controlled Trial (no. ISRCTN 22488978).

Summary

Ovarian cancer is the fourth most common cause of cancer death in the UK and the five-year survival rate is significantly less than in some European countries. Whilst the reasons for this poor performance can only be speculated upon, it is clear that early diagnosis significantly improves the outcome for women with the disease. This has created considerable interest in the potential of ovarian cancer screening. For this reason UKCTOCS was funded to assess the impact of screening on mortality from ovarian cancer. TVS scanning is core to such screening either as a first or second-line test.

Increasing understanding of the development of ovarian malignancy, the potential of annual screening for ovarian cancer has been shown through models which suggest that the window of opportunity to save life after the detection of an early stage serious tumour is more than four years. However, the model also indicates that the sensitivity of such screening is dependent on detecting tumours whilst they are still very small. This would be a significant challenge for sonographers performing TVS who usually carry out examinations for diagnostic purposes on women presenting due to symptoms arising from a gross tumours that have developed to late stage. Indeed, the lack of mortality benefit found in the ultrasound arm (USS) could be explained by the higher level of performance needed when TVS is given as an annual screen for ovarian cancer rather than as a diagnostic. However, there is a lack of information about how this challenge might be addressed as ovarian cancer screening has only been performed in a few clinical trials. In this respect the results of the TVS screening by UKCTOCS has some potential to improve understanding of the issues that need to be addressed, especially in respect of Quality Control (QC).

This thesis aims to describe some of the difficulties associated with implementing effective QC for TVS examinations used for the screening of asymptomatic menopausal women in a population. It also suggests some solutions to these problems in terms of the tools detailed in Chapter 6 that serve to automate the collection of independent and objective quality metrics for such examinations (see also the studies in Chapters 7, 8 and 9). Combining these sorts of reliable QC metrics with the type of Quality Improvement (QI) programme discussed in Chapter

3 creates the possibility of advancing the performance of TVS screening towards the level required for an effective ovarian cancer screening test. However, achieving the sensitivity and specificity required for primary ultrasound screening to provide real mortality benefit in any future national ovarian cancer screening programme is likely to remain an outstanding issue unless improvements in ultrasound technology (Chapter 2) allow the development of novel biomarkers. Therefore it might be that ovarian cancer screening remains dependent on tests based on blood serum protein biomarkers with TVS used as a second line diagnostic test much like the screening approach of the multimodal (MMS) arm of UKCTOCS.

Chapter 2: Role of Technology in Quality of TVS Scanning

Introduction

This chapter considers the role of technology in achieving high quality TVS scanning. It describes the principles of ultrasound and explains how the various components of an ultrasound machine are combined to produce high quality images. It also reviews the ultrasound equipment used by UKCTOCS, its selection criteria and what was considered important in terms of delivering high quality TVS scanning. The chapter concludes with a brief discussion about how advances in technology might impact the equipment used for TVS scanning in any future large-scale screening programme for ovarian cancer.

Ultrasound Principles

High frequency sound waves are termed ultrasound and have numerous applications including medical diagnostic imaging. High frequencies have smaller wavelength so allow better spatial resolution for a given aperture, as specified by the Rayleigh criterion³¹. However, when passing through typical body tissue the signal suffers attenuation of about 1 dB/cm/MHz³², so higher frequencies also result in poorer penetration. Therefore, frequencies are chosen to optimise resolution at the depth of the tissue being viewed. Most medical diagnostic imaging uses frequencies in the range 2-20 MHz, with frequencies between 8-13 MHz being typical in transvaginal sonographic (TVS) examinations.

Absorption and Reflection of Ultrasound

The use of ultrasound for diagnostic imaging follows principles that are similar to those employed in radar or sonar equipment. A transducer probe periodically transmits a pulse of ultrasound and then switches to receive mode in order to detect the reflections of this beam as it passes through body tissue. The received signal comes from reflections in tissue at various distances from the transducer, commonly categorised in terms of the near and far field.

The received signal depends upon the changes in speed and absorption that a sound wave suffers as it passes through different materials. For example, there will be a large reflection when the sound wave arrives at a junction between bone and fat. Conversely, there will be no reflection whilst the sound wave travels through material that impedes sound waves uniformly – i.e. materials with same acoustic impedance. The time taken for a sound wave to be reflected back to the transducer depends upon the acoustic impedance of the material as well as the amount of the material it must travel through before being reflected. Consequently, the reflection of a junction between bone and fat that lies deep in the body will arrive after a similar junction close to the surface.

Resolution

The resolution of ultrasound is a measure of the equipment's ability to separate points in tissue in respect of space, time or type. Spatial resolution refers to

separation in terms of axial and lateral directions. For example, in a section of an ovary the axial resolution determines whether in the axis of the beam the epithelium might be seen as separate to the stroma, whilst the lateral resolution determines whether at 90° to the beam a small septum might be seen suspended between the walls of a cyst. Similarly, temporal resolution describes whether the movement of two points over time can be seen and contrast resolution describes the ability of the equipment to separate tissue at different locations with similar acoustic impedance.

Modes of Operation

The simplest form of ultrasound machine is called an A-mode. It plots the time reflections are received on the X-axis and the amplitude of the reflection on the Y-axis, so that large peaks on the graph show materials with large differences in acoustic impedance and the distance between these peaks shows their respective depths. The limitation of the A-mode device is that it can only show features in a single axis with respect to the ultrasound beam.

B-mode machines operate in a similar way to A-mode but move the transducer at a constant speed through an arc (or line) to show features in a plane. This allows the features to be displayed in a two dimensional (2D) image with the Y-axis corresponding to their depth and the X-axis corresponding to the scanning plane. The amplitude of the signals received by the transducer from each point in the resultant image plane is indicated by the intensity of the pixels in order to produce a grey-scale picture. Areas of the image showing tissue that is homogeneous in terms of acoustic impedance are displayed in black and are termed hypoechoic (or anechoic) because they cause no reflection of the ultrasound waves. Conversely, tissue that causes strong reflection is displayed in white and termed hyperechoic.

Ultrasound machines may be operated in a variety of other modes, each with a particular purpose. For example, F-mode provides an additional colour overlay for the grey-scale image showing blood flow in the tissue. This is a representation of the frequency shift between the transmitted and received ultrasound signal resulting from the Doppler effect such that blood moving towards the probe had a positive Doppler shift and is shown in red, whereas blood flowing away from the probe has a negative shift and is shown in blue. F-mode (Doppler) imaging has diagnostic value

in TVS ultrasound examinations as abnormal blood flow in an ovary is often a sign of malignancy. Doppler effects are also employed in other ultrasound modes such as D-mode (Spectral or Pulse Wave) which is used to obtain blood flow velocity in a small region.

The assembly of a collection of B-mode images along an axis which is orthogonal to their 2D plane allows the production of a three dimensional (3D) image. A further dimension may be added by sequencing the images by time (4D), though most modern machines display B-mode images in real-time during TVS examination to give the same movie effect with frame rates typically greater than 15 per second. Overlays containing supplementary information like blood flow are also considered as providing additional dimensions to the image, so machines may often be described as being 5D or 6D.

Beam Focusing

The ultrasound beam has a cylindrical shape in the near field, but extends into a cone shape in the far field. Its actual dimensions depend upon factors such as the transducer design, frequency and the focusing applied. The dimension of the beam in the scanning plane determines the beam width, while the dimension in the orthogonal plane determines the slice thickness. The intensity of the beam (power) is not uniform throughout its length or section resulting in a given feature having stronger echoes at certain distances from the face of the probe than others, as well as echoes being greater at the centre of the beam than at the edges. Consequently, a better view of the feature can be obtained by focusing the beam in both transmit and receive mode to create one or more focal zones where the intensity is concentrated into a small area to give a more powerful echo which in turn maximises the spatial resolution.

Focusing may be achieved both by a structural lens attached to the transducer and by electronic control of the transmitted and received signal. Electronic focusing exploits the interference effect that arises when there are multiple sources of a signal. This is the same type of interference seen when two stones are thrown into a pond; they create a series of waves which reinforce each other at points where their maxima meet, whilst diminishing themselves where maxima and minima coincide. By

adjusting the spacing and delay between the signal sources the interference pattern can be manipulated to amplify the signal at one or more focal points for a given signal frequency. This effect is used in phase array probes.

There is a trade-off between the size of the focal zone and resolution such that higher resolution results in a smaller focal zone. Therefore, the operator optimises the image by controlling the focal point as well as the number and size of the focal zone(s). However, it should be remembered that the depth of field also depends on the frequency of the ultrasound signal, with higher frequencies having poorer penetration. Therefore, the high resolution that results from a 20MHz signal comes at the cost of a focal zone that has application only for near field targets such as dermatology.

Specification of Ultrasound Machine Resolution

There are no published standards giving quantitative measures of imaging performance for machines used to perform TVS examinations. Therefore it is assumed that a machine with adequate performance will be used for such scanning and selection is made on the basis of subjective evaluation assisted by reviews published on reseller websites like GPS Medical³³.

The lack of a more structured approach to comparing ultrasound machines was recognised by Wynd et al³⁴ in their 2009 paper. The results reported for image quality comparisons were based on subjective assessments by 16 physicians of seven industry representative machines in terms of a 10 point scale ranging from poor to exceptional quality. The mean score of achieved by all machines was 6.8 (range 5 to 9, SD 1.52) and there was a statistically significant difference between the machines in terms of reviewer score ($P < 0.001$). It was shown that purchasing decisions for ultrasound machines were not just based on image quality, since 56% of respondents chose a machine with image quality score six whilst only 13% selected the machine with an image quality score of nine (highest). The results suggest that while image quality does vary between machines, it is subjective and machines have good or better performance in this regard irrespective of price in the range \$25K to \$150K.

In an attempt to provide a more quantitative approach to specifying medical ultrasound machine resolution, Moran et al³⁵ have recently suggested use of the

Resolution Integral as an objective measure of ultrasound machine imaging performance. This is based on earlier work by Pye and Ellis³⁶ who assumed that the “two key characteristics of high quality ultrasound images are narrow beam width and good penetration into soft tissue”. In the case of weakly focused imaging beams they proposed that a metric to provide a quantitative value of the machine’s spatial resolution – i.e. its ability to visualize two separate points in tissue that lie in close proximity. It addresses the difficulty of providing a single value that combines resolutions in both the axial and lateral direction of the ultrasound beam which due to the nature of the focal zone is typically tapered in the middle rather than being shaped like a uniform column. This is achieved by integrating resolution over the axial direction of the beam in order to give a more accurate value than might be obtained by simple calculation based on the assumption that the beam column has a uniform width along its entire length; see Figure 2-1. The Resolution Integral has potential in allowing performance comparisons to be made between different machines used to scan the same Tissue Mimicking Material (TMM).

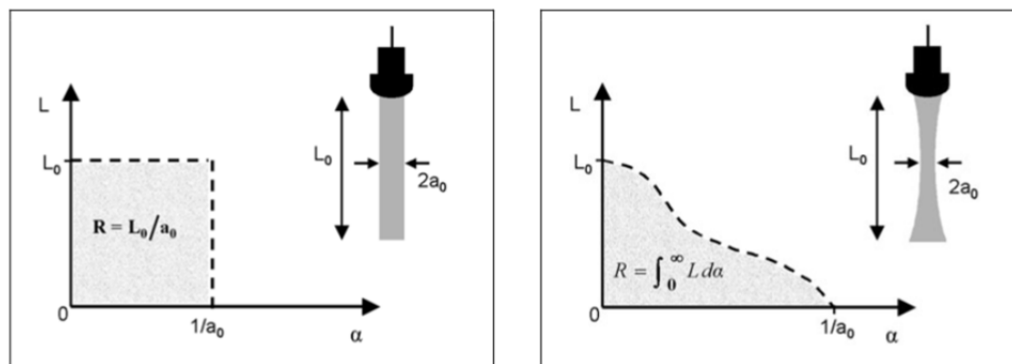


Figure 2-1: Resolution Integral Calculation^a The left side shows the resolution integral R for a collimated beam with penetration L_0 and beam width $2a_0$. The right side shows the resolution integral R for a weakly focused beam with penetration L_0 and beam width $2a_0$. © IOP Publishing. Reproduced with permission. All rights reserved.

To demonstrate the method Pye and Ellis constructed a novel test object from a block of TMM and measurements of resolution were made at different depths using a variety of machines and probes. The values generated by the Resolution Integral provided good discrimination between the various types of equipment and an

^a Pye, S. D., W. Ellis, and T. MacGillivray. 2004. ‘Medical Ultrasound: A New Metric of Performance for Grey-Scale Imaging’. *Journal of Physics: Conference Series* 1 (1): 187. doi:10.1088/1742-6596/1/1/041

association was found between its value and subjective assessment of their clinical performance.

Moran et al reviewed the work of Pye and Ellis as well as others experimenting with the Resolution Integral and showed how the metric reflected improvements in image technology made over 20 years by measuring the performance of a number of machines purchased over this period. Therefore it was concluded that the Resolution Integral had potential as an objective measure of imaging performance. However, at present this metric is not commonly specified for ultrasound machines so its use in equipment selection is limited.

Challenges of Image Interpretation

In addition to reflections useful for clinical diagnosis, the received signal also contains components resulting from the scattering, refraction and divergence of the main ultrasound beam. Such signals may change the way objects appear or create objects in the display which do not correspond to actual structures in the body. Variation in beam intensity by length and section may also make features appear differently in the image to their actual form in the body. The challenge of image interpretation is further complicated by objects failing to be displayed due to acoustic shadowing – i.e. a feature of high acoustic impedance behind a feature of low impedance causing its signal to be masked. Moreover, unwanted signals may also be received from side lobes which can appear at various angles to the main beam lobe which is directed at 90° to the plane of the transducer face. Such signals can introduce artefacts into the displayed image which can also make its evaluation difficult.

The consequence of such artefacts and distortions in the ultrasound signal is that not only could objects in the body fail to be imaged, but phantom objects may also appear and objects in the display may be completely transformed in terms of depth or size. Therefore the correct interpretation of an ultrasound image requires a sonographer to have considerable skill and experience. The development and assessment of these skills is discussed in Chapter 3 under the heading “Quality Assurance of Sonographers”.

Safety

Most side-effects of a TVS exam result from probe insertion and manipulation or from poor infection control, such as inadequate sterilisation of the probe between examinations. However, in other respects diagnostic ultrasound is considered a very safe imaging modality, though it does apply energy to biological tissue so has the potential to cause harm if misused³⁷. The damage can result from thermal or non-thermal effects. Thermal effects result from the transfer of energy when tissue absorbs the ultrasound waves. The amount of energy transferred depends upon the type of tissue, the beam frequency and power as well as the cross-sectional area and exposure time. Thermal effects are most significant for high frequency beams going through tissue with a high absorption coefficient which is unable to dissipate heat. Accordingly, the value of the thermal index (TI) is displayed on the ultrasound machine's monitor alongside the real-time scanning image. This corresponds to the ability of the beam at the current power setting to raise the tissue temperature by 1 degree centigrade at the focus point.

The main non-thermal effect is cavitation defined as the formation, growth, movement and rupture of micro-bubbles in tissue. Cavitation is more pronounced when contrast agents are introduced into tissue in order to facilitate some types of examination. However, it is usually only considered a problem when the intensity of the beam is high enough to result in the rupture of micro-bubbles creating localised high pressures and temperatures in tissue. In order to avoid such problems, the value of the mechanical index (MI) is displayed on the ultrasound machine's monitor. This corresponds to the peak negative pressure divided by the square root of the beam frequency. Guidelines for maximum values of MI (and TI) are given by bodies, such as British Medical Ultrasound Society (BMUS). However, the normal settings for TVS examinations should not approach these thresholds.

Review of Imaging Technologies used for Cancer Screening

In order to understand why ultrasound is the primary choice for initial ovarian cancer screening it is useful to consider other modalities commonly used for such medical imaging, such as:

- **X-ray radiography** – placing the body between an x-ray source and a detection plate creates a projection of its internal structures, or may be used in conjunction with contrast agents (e.g. barium) to allow real-time fluoroscopy. Whilst the costs of obtaining and analysing an x-ray image are typically greater than performing an ultrasound scan, this type of imaging can still be performed at low cost. It is considered a reliable means of diagnosing various forms of cancer including breast (mammograms) and lung (chest radiographs). The robustness of interpretation by an oncologist depends on the image obtained by the technician and in this respect ultrasound is more effective as a diagnostic approach than x-ray because it allows the sonographer to explore body tissue in order find abnormalities. The radiation dose required for a single x-ray is low, but can become significant when large numbers of images are required over a short period. There is also the possibility that such x-ray examination might increase the frequency of radiation induced mutations in the population if used in a national screening programme, though as yet there is no conclusive evidence of this effect. By comparison ultrasound is safer as within normal operating ranges there is no associated risk of cellular damage.
- **X ray computed tomography (CT)** – alternatively termed Computed Axial Tomography (CAT), a tomographic image is produced by assembling under computer control a series of two dimensional image slices which are generated by spinning a beam of x-rays around the body and detecting differential absorption by tissue from multiple angles. CT equipment is more complex than a standard x-ray machine which results in higher capital and maintenance costs. It also requires more sophisticated processing to produce the images needed for interpretation which again adds to the cost. However, it is considered a reliable means of diagnosing various forms of cancer. The interpretation of the images by an oncologist is less dependent on the work of

the technician performing the scan, but as a diagnostic approach it lacks the facility for exploration that characterised ultrasound. The radiation dose associated with CT scanning is necessarily higher than for a single x-ray which makes it even less attractive as a screening approach than ultrasound in terms of patient safety.

- **Magnetic Resonance Imaging (MRI)** – originally called nuclear magnetic resonance (NMR) imaging, it applies powerful radio frequency (RF) magnetic fields to polarise and excite the hydrogen nuclei of water molecules in tissue and then detects the resultant spatially encoded signal in order to create a two dimensional slice of the body. Assembling these slices into a tomographic image allows a three dimensional view to be created with high resolution and soft-tissue contrast. The generation of the required magnetic fields requires large and complex equipment which is expensive in terms of both capital and running costs. This makes MRI an expensive imaging modality which precludes its use as a primary screening modality. However, MRI has been found to be a more reliable means of diagnosing a wider range of cancers. This is because the differences between abnormal and normal tissue are often much clearer than in other imaging modalities and for this reason MRI images are considered easier to interpret and result in a more certain diagnosis. In addition MRI does not apply ionizing radiation to the body and there are no known side-effects to its strong magnetic fields, unless the patient has metal embedded in her body³⁸.
- **Positron Emission Tomography (PET)** – radioactive isotopes are administered to the patient in a form that allows them to be absorbed and detected as markers for metabolic utilization in order to show rapidly growing tissue such as tumours. Therefore PET provides a further dimension for images from other modalities like MRI or CT to show metabolic activity in respect of spatial position for a wide range of cancers. The cost of administering these isotopes is high (£1,000 to £2,000 per dose) and additional processing is required to provide images in addition dimensions. However, the information provided by these dimensions can result in patients receiving a more assured screen. The half-life of the radioactive isotope is

short and the committed dose is low. Therefore the risk of resultant cellular damage is low, but still present.

Ultrasound is unlike the above modalities in that the value lies in the examination, not in the images it produces. During a TVS examination the only images routinely preserved are the static images; screen-shots of the real-time display made to record specific detail or allow measurement of features such as ovary size. In the NHS these images will seldom be viewed after the examination except when an abnormality is found. This is very different to the situation in other imaging modalities in which image acquisition is performed by a technician and subsequent interpretation of the images is performed by a qualified doctor, typically a radiologist or consultant with relevant expertise.

When performed by a highly skilled expert a TVS exam is a relatively inexpensive, convenient and very safe way of checking for abnormalities associated with ovarian cancer. It provides a real-time view of internal organs and blood flow which can allow discrimination between benign and malignant ovarian masses as well as between ovarian and non-ovarian masses³⁹. In order to obtain this outcome the examiner creates a mental three dimensional map of the internal structure of the Adnexa from the two dimensional image shown on the main display. Experienced TVS examiners are so skilled in this task that they seldom use the machine to generate three dimensional images. The advantage of this approach is that the examiner can target particular structures like the ovaries and then optimise the image by adjusting the controls as well as externally manipulating the patient's body to move organs such as the bowel which may be obstructing the view. This gives ultrasound much more flexibility than modalities like MRI as the examiner plays an active part in determining what is visualised.

The main disadvantage of ultrasound as a diagnostic imaging modality is that assuring a consistent level of quality for examinations is challenging. This is why the quality control of ultrasound images is an important topic, as discussed in subsequent chapters. However, before considering the quality of such images it is helpful to understand how they are produced.

Technical description of an ultrasound machine

The main components of a typical ultrasound machine used for performing TVS exams are shown in Figure 2-2³².

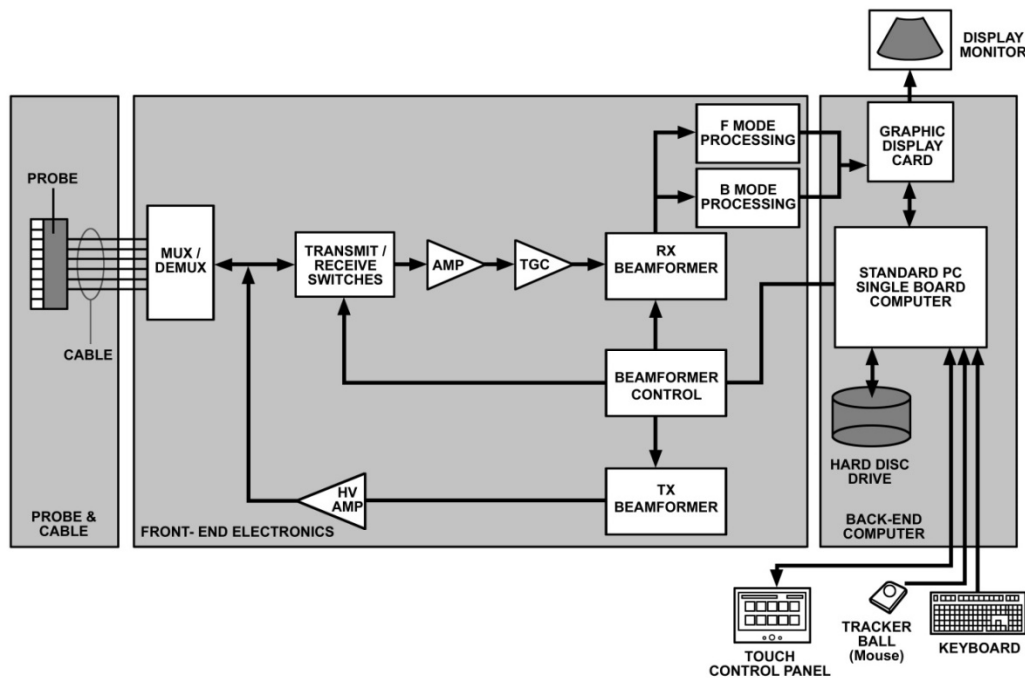


Figure 2-2: Medical Ultrasound System Block Diagram. There are three subsystems; the probe & cable, front-end electronics and the back-end computer. The front-end electronics switches between transmit and receive mode such that the ultrasound signal is generated and then its reflections are detected and processed before being displayed.

Probes and cables

An ultrasound probe usually contains backing material, a piezoelectric transducer, an acoustic lens and a matching layer as shown in Figure 2-3. The matching layer aims to reduce the acoustic mismatch that occurs as ultrasound travels from the surface of the transducer through the probe lens assembly into the tissue under examination. Gel is also applied between the surface of the probe and the patient's skin to further reduce signal loss due to acoustic mismatch.

The physical shape of the piezoelectric transducer and the type of backing material influence the probe's quality (Q) factor. A high Q means that the energy of the excitation pulse only slowly dissipates so predicating the use of short duration

pulses. However, short pulses need higher voltages to deliver the same amount of energy as a long pulse. It is for this reason that medical ultrasound transducers operating at high frequencies are constructed to have a high Q with high voltage pulses of short duration used to optimise the axial resolution. Typically, the transducer is operated at 500 volts with a spatial pulse length (SPL) of 0.6 μ S which allows two or three cycles of a 5MHz primary harmonic frequency.

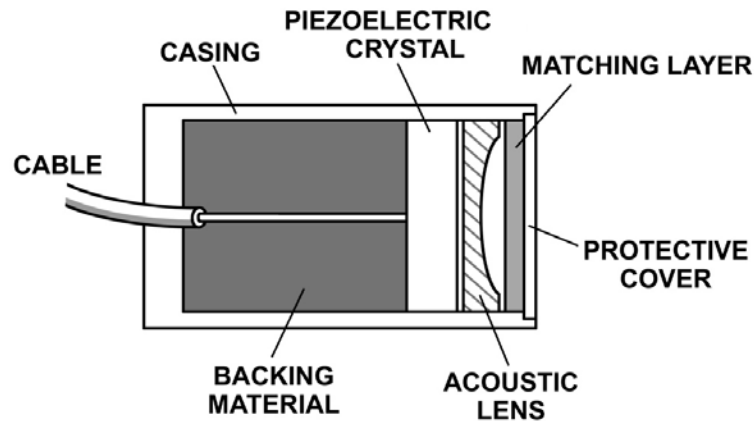


Figure 2-3: Main Components of an Ultrasound Probe. The piezoelectric crystal is responsible for both transmitting the ultrasound signal and receiving its reflections. Whilst the actual design of a probe used for TVS is considerably more complex, it contains these same basic components.

A modern ultrasound machine is operated in B-mode by assembling a collection of piezoelectric transducers into an array each with their own lens. This avoids the need to move a single transducer in relation to the object being imaged in order to sweep across the field of view. In a simple stepped array, the elements are activated in sequence with the shape of the probe head determining whether the scan is linear (straight), arc (concave) or sector (convex). However, in the more complex phase array all the elements are simultaneously activated and the beam is steered across the field of view by adjusting the delay profile of the individual array elements; a technique called beamforming (Figure 2-4). Consequently, the phase array probe can create a scan without requiring the head to be a particular shape and size.

The EC4-9IS probe used in UKCTOCS TVS examinations after January 2008 is a phased array type, specifically modified for endocavity use (similar to the probe used before this date). Accordingly, it has a small tip containing 192 elements arranged

into an arc giving a sector scan with 150° field of view. The grouping of elements in an array so that they fire at the same time allows the power of the ultrasound beam to be increased. However, in much the same way that increasing the aperture of a camera lens reduces depth of field, increasing the number of elements in each group results in a decrease in the range of focus. Therefore, the challenge when designing an ultrasound probe is to pack as many transducer elements into the space available for the array whilst maintaining individual performance and control.

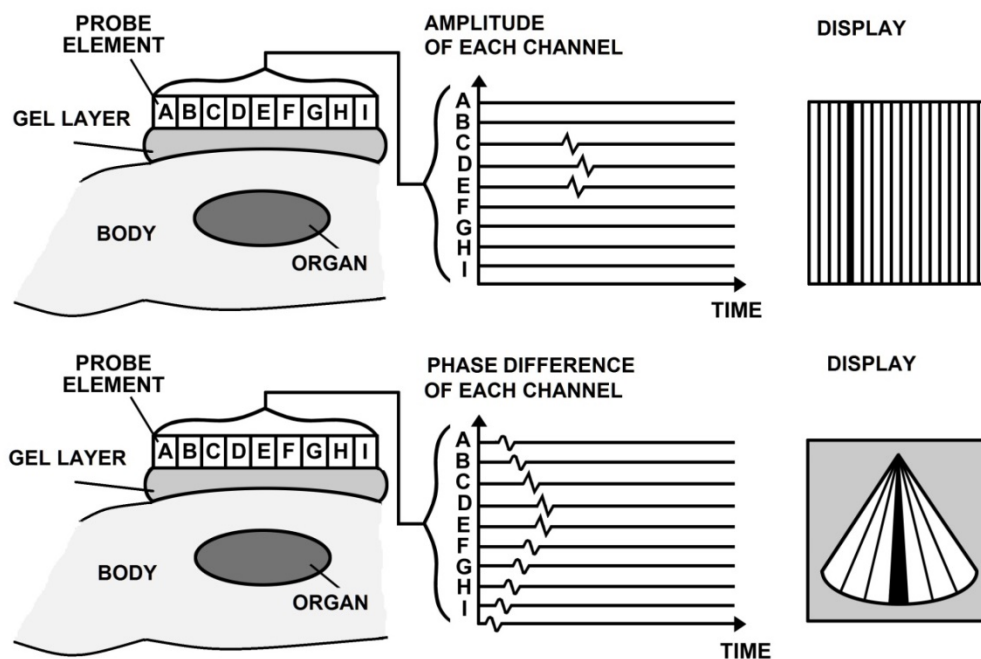


Figure 2-4: Stepped and Phase Array Signal Formation. A linear stepped array (top) forms a beam line from the signal received by each probe element activated in sequence during a scan frame. The phased array (bottom) forms beam lines whose direction is determined by the delay profile of the pulses received from all the probe elements which are simultaneously activated during a scan frame.

In receive mode the transducer needs to be able to detect signals with a significant range of values from the large signals from reflections in the near field to the small signals from the far field. This requires electronics with a wide dynamic range in order that adequate contrast resolution can be provided for the entire viewing range. The ratio of the smallest measurable signal to the largest can be greater than 300,000,000:1 (170dB) and the processing of the very small signals in this range presents a further challenge in terms of separating the signal from the noise that is always present in electronic circuits. Therefore the components in the transducer

must not only have a wide dynamic range, but also minimise noise in order to present a signal to the front-end electronics with a suitable signal-to-noise ratio (SNR).

The challenge of achieving an acceptable signal to noise ratio over the wide dynamic range associated with ultrasound continues to be addressed by advances in semiconductors and circuit design with resultant improvements in image contrast and sensitivity. However, progress is ultimately impeded by the requirement to drive a transducer with the complex high voltage signals needed for ultrasound transmission and then switch to the low voltage signals received from ultrasound echoes. It is for these reasons that the digitalisation of the signal at the probe has proved so difficult. Accordingly, improvements in image contrast and sensitivity are limited by the need to transmit low level analogue signal from the probe to the front-end electronics via a two meter long cable.

The cable is often one of the most expensive items in the parts list of an ultrasound machine. It is typically composed of a large number (48 – 256) of micro-coaxial cables bound together in an outer sheath of material to facilitate sterilisation and provide mechanical support. Each micro-coaxial cable connects the ultrasound transducer to a channel in the front-end electronics. The cable is a source of significant signal loss and so impacts the signal to noise ratio which is an important parameter of system performance. Therefore it is important to ensure that there is a good match in terms of capacitance and inductance between the transducer circuitry, the cable and the front-end electronics.

Front-end electronics

The key parts of the front-end electronics are the transmit (Tx) and receive (Rx) beamformer circuits which are concerned with the generation of the ultrasound signal and the reception of its reflections in order to optimise the display of tissue at the desired focal point. These circuits are typically concerned with analogue signal processing, so particular attention must be paid to reducing noise and distortion which will limit overall system performance.

The beamforming electronics are responsible for controlling an array of ultrasound transducers so the spacing and delay of the signal can be adjusted to put the focal

point(s) under operator control, whilst also allowing for the greater energy needed to penetrate deeper into tissue and compensating for the relative attenuation of the return signal. The Tx beamformer determines the timing and pulse shape of the signal sent to each element of the transducer array. In some machines the output of the beamformer is analogue, but in others better control of the signal is achieved by outputting a digital signal which is converted into analogue by a dedicated DAC (digital to analogue converter) chip. The signal is then amplified and converted into the high voltage (HV) required to drive the ultrasound transducer when operating in transmit mode. Often a de-multiplexor circuit connects the output of the amplifier to the appropriate transducer, though more advanced machines have separate DACs and HV amplifiers for each transducer in the array allowing them to be individually controlled in terms of phase and amplitude.

A de-multiplexor (DEMUX) allows a signal from a common circuit to be directed to multiple channels and conversely a multiplexor (MUX) allows signals from multiple channels to be handled by a single processing circuit. This reduces the cost and complexity of the front-end electronics by avoiding the need for each element in the array of ultrasound transducers to have its own separate signal processing circuit. However, multiplexing multiple channels into one processing circuit limits flexibility as such processing cannot then be made channel specific.

Once the ultrasound signal has been transmitted by the transducer the T/R switch in the front-end electronics switches into receive mode, so the analogue signal from the transducer can be amplified and processed by the front-end electronics. This T/R switch is necessary to protect the input of the low noise amplifier (LNA) from the high voltage pulses going to the transducer in transmit mode and is typically implemented by a series of bridge diodes. The LNA needs to be able to handle the large dynamic range of signal that results from the very strong echoes of near-field objects as well as from the very weak echoes of far-field objects much deeper in the body. This results in a design trade-off between a circuit able to handle a large signal and a circuit able to handle a very low signal without adding noise.

The operation of the machine in B-mode requires the implementation of time gain compensation (TGC) circuit which allows more amplification to be applied to signals arriving later in the cycle in order to compensate for the attenuation they

suffer as a consequence of coming from reflections deeper in the body. This is achieved by using a variable gain amplifier (VGA) to process the LNA output signal. In an Analogue Beam Forming (ABF) circuit the output of the VGA circuit is fed to a variable delay line before being summed and converted into a digital signal by a dedicated ADC (analogue-to-digital converter) chip. However, in a digital beamforming circuit (DBF) the output of the VGA is first digitised by an ADC and then fed to a first-in-first-out (FIFO) buffer to apply the required delay before being summed by a digital adder; see Figure 2-5.

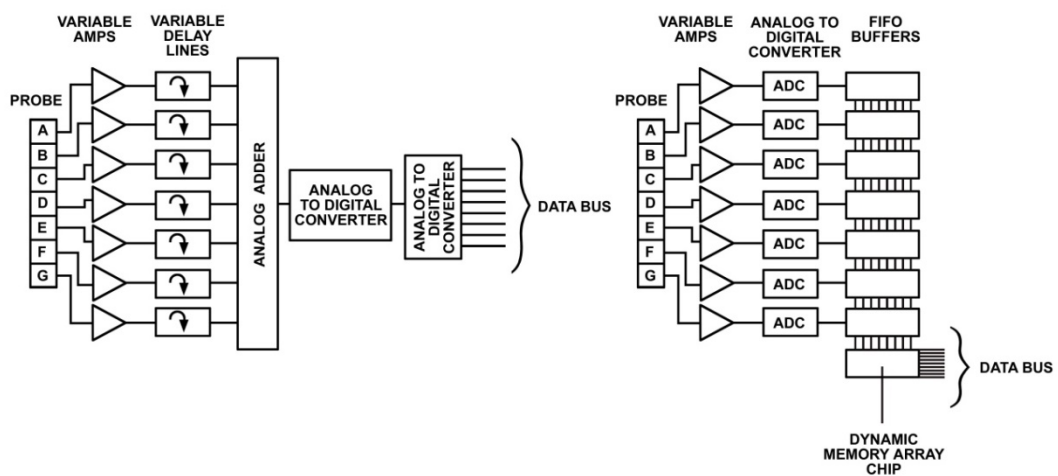


Figure 2-5: Comparison of Analogue and Digital Beamforming. The analogue circuit (left) uses variable delay lines to store signals from the probe channels which are aggregated by the adder circuit before being converted to a digital value for the data bus. The digital circuit (right) converts the signals from the probe channels into digital values which are stored in a first-in first-out buffer memory. The contents of these buffers are then sampled by the DMA chip to aggregate the signals received by each channel at selected delays and so provide a value for the data bus.

Whether the beamforming is analogue or digital, it typically results in a digital value corresponding to the amplitude of the ultrasound signal received by the array of transducers as reflected by tissue at the given focus point in the field of view. The location of this point in the field of view is set by changing the relative delay of the signal from the different elements of the array. Therefore a two dimensional (2D) picture can be produced by adjusting the delay for each channel in order to scan the field of view progressively in the X and Y planes (raster scanning). The addition of scanning in the Z plane allows the production of a three dimensional (3D) picture

and by periodically repeating the scan a series of frames can be generated showing the field of view in the fourth dimension (time).

The beamformer circuits are controlled by dedicated processors (see Appendix I) that take responsibility for directing the raster scanning needed to form images as well as synchronising the associated transmission of pulses. Field-programmable gate arrays (FPGA) are often used in this context as they offer better input/output performance than microcontrollers and can be developed more cheaply and flexibly than application-specific integrated circuits (ASIC). In addition, Digital Signal Processor (DSP) chips may be employed to process the Rx beamformer signals extracted from the FIFO buffer at specified delay points. This allows the image to be built-up in the display memory and by storing successive frames a replay facility may be implemented.

Developments in the semiconductor industry have allowed ultrasound machine manufacturers to integrate very powerful processors into their machines which have transformed the front-end electronics, typically implemented on circuit boards plugging into a common data bus. However, the major advances result from advances in software capability allowing rapid prototyping on hardware simulators, hardware languages for DSPs and so forth. Such developments have transformed front-end electronics development making it faster, cheaper and more effective. This led to the step-change in ultrasound machine capability that occurred in the 1990s and continues to drive incremental improvements in these machines to this day. The impact of improvements in data compression, digital filtering and the changing nature of the signal processing pipeline has particular relevance to machines used for TVS scanning.

Data Compression

Transducers with elements arranged in a matrix provide signals that facilitate artefact removal as well as the volumetric scanning needed to produce three dimensional images. However, processing the phase as well as amplitude of the radio frequency (RF) data arriving from a large number of elements in these types of probes is a significant challenge. The problem is compounded if it is necessary to store or transmit the data to a remote location for future analysis. The front-end electronics can facilitate matters by compressing the RF data whilst retaining the components

needed to recover the small and transient features important for image rendering and interpretation. This requires a combination of high performance hardware and efficient software. Graphics Processing Unit (GPU) hardware employing hundreds of processing cores, such as produced by NVIDIA, is particularly suitable to this task, whilst algorithms such as those reported by Govindan and Saniie⁴⁰ can provide the basis of the software implementation.

Digital Filtering

In addition to compression, the signal will also be processed by the front-end electronics in other ways, for example to support the filtering required by harmonic imaging. The development of tissue harmonic imaging has led to better depth of field, less artefacts and enhanced contrast. It works by filtering the received signal so the image is constructed from the harmonics rather than the probe's fundamental frequency. This gives the deeper tissue penetration associated with the transmitted lower fundamental frequency (i.e. 3MHz) and at the same time the better resolution associated with the received higher harmonic frequencies (say 6MHz, 9 MHz, etc.). In addition to improving depth of field by giving better lateral resolution at given depth, the technique also reduces near field artefacts as harmonic signals from tissue only become significant in the far field where the energy of the transmitted ultrasound beam is focused. However, its greatest advantage is the increase in contrast resulting from the reflected harmonic signal's reduced dynamic range which allows much greater segregation of the greyscale amongst tissues with only small differences in acoustic impedances. Tissue harmonic imaging depends on high performance transducers with good SNR as well as highly capable Digital Signal Processor (DSP) chips and efficient software able to perform the necessary digital filtering.

Changing the nature of the signal processing pipeline

The advent over the last five years of new approaches to graphics processor design combined with interfaces that allow them to undertake general processing tasks using software written in languages like C++ has resulted in some manufacturers re-engineering the architecture of their front-end electronics, extending it well beyond its traditional role in beamforming. In addition to performing data compression and filtering (described above), the front-end electronics may also take on traditional

back-end computing tasks such as the formation of B-mode images. Siemens Acuson SC2000 uses the NVIDIA Quadro2000 GPU card with the CUDA software interface to implement the signal processing pipeline shown in Figure 2-6.

It can be seen that the GPU card is involved in the acoustic domain processing of the digital signal from the phase array beamformer circuits in another front-end electronics card. The resulting compressed data is held in cine memory which is arranged in a way that facilitates access by software also running on the GPU card which is responsible reconstructing the image and then performing final image enhancement, such as speckle reduction, before rendering it as the 2 dimensional scan in the display memory. It can be appreciated the development of such architecture makes it difficult to identify any physical interface where signal processing by the front-end electronics stops and where back-end processing starts.

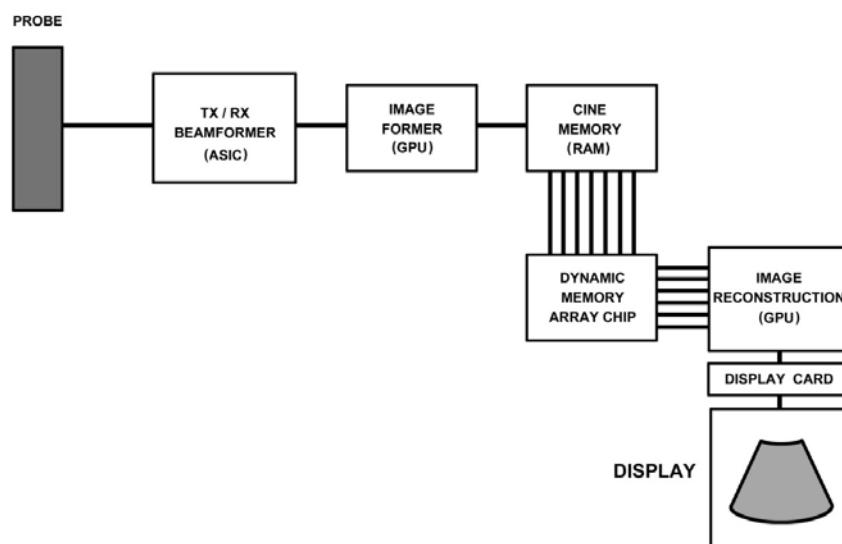


Figure 2-6: Modern Ultrasound Signal Processing. Beamforming is performed by an Application Specific Integrated Circuit (ASIC) which provides the raw signal for image formation in a Graphics Processor Unit (GPU) based subsystem. The Image Former feeds the Random Access Memory (RAM) which is sampled by the DMA chip so the image for display can be constructed by the image reconstruction subsystem. Appendix I gives further details of these processor types.

Cine Memory

Cine memory is a feature of most modern machines and is essentially a buffer for storing data that can be reconstituted as a series of image frames. This allows the operator to rewind through frames shown on the main display, much like the rewind

facility on a Cable TV box. It also permits the processing of data frames in order to improve the quality of the displayed image. For example, despeckle filtering^{41,42} to remove the light and dark spots in the image caused by inference effects arising from ultrasound reflected by small structures (small or equal to the wavelength) in otherwise homogeneous tissue. Alternatively, the spatial resolution of ultrasound images may be improved beyond the limits imposed by Rayleigh diffraction by processing a collection of frames to give a super-resolution image^{43,44}.

HistoScanning

Whilst it may be helpful for the ultrasound machine to remove speckle when viewing anatomic features like ovarian cysts, speckle can yield useful information about microstructures in tissue. This technique is called HistoScanning⁴⁵ and reflects the fact that speckle arises from the stream of small echoes from microscopic boundaries associated with the histological structure of tissue. HistoScanning is commonly used to aid diagnosis of prostate cancer, but it has also been used by Lucidarne et al⁴⁶ in order to differentiate malignant ovarian masses from non-malignant lesions.

Although it is possible to perform this type of analysis by post-processing voxels of 3D image data obtained from a TVS examination, a more typical approach involves real-time processing of the raw RF data. This allows the extraction of much more information about the back scattering of the ultrasound signal. It also allows dedicated processors in the front-end electronics to classify the results so areas of potential malignancy can be highlighted in the display image.

Back-end Computer

The back-end computer is responsible for providing the user interface which ultimately controls the beamforming circuits (Tx and Rx). It also runs a number of other software programs that allow the operator to undertake tasks like taking measurements and recording images from the front-end electronics. Typically a standard commercial computer is used for this purpose. In this way ultrasound machine manufacturers can concentrate on developing the specialist hardware and software needed to provide their unique functionality whilst utilising off-the-shelf products for generic tasks such as display, disk storage, network communication and so forth.

Many ultrasound machines have their back-end computer implemented on an IBM PC compatible single board computer. Typically, its operating system will be provided by some version of Microsoft Windows which, whilst multi-tasking, is not a real-time operating system. This is because it does not provide any guarantees in terms of interrupt latency or thread switching, so its processes may not run consistently. For this reason, Windows processes cannot be expected to attempt the sort of signal processing performed by the front-end electronics (see above). Therefore, the back-end computer is involved with tasks that are not particularly time critical.

The ultrasound machine used by UKCTOCS for the last part of the trial is the Medison Accuvix XQ, as mentioned later in the chapter. It's typical of such an off-the-shelf approach to back-end architecture. The hardware is comprised of an IBM PC compatible main board containing a Pentium IV 2.8GHz processor and 2GB RAM and its operating system is Windows 2000⁴⁷. The main board computer plugs into a common bus shared by the display board as well as custom boards which implement the front-end electronics and connect to the probe by a cable. In this way the machine is based on a modular chassis which is easy to enhance. The software programs run by the Medison Accuvix XQ on its back-end computer are typical of those provided by most modern ultrasound machines:

- **Main control:** responsible for providing the user interface to the machine operator (Sonographer). This includes the main display, standard user controls such as keyboard and tracker ball as well as a variety of custom controls for specific ultrasound functions (see below), machine setup, etc. The main control program is managed using a touch-screen module which allows selection of machine operating mode, display mode and so forth.
- **Ultrasound image viewing:** responsible for displaying the ultrasound image as rendered in the display memory in the mode selected by the machine operator. This may be the real-time image, images stored in memory (cine) or images stored on the machine's hard-disk.
- **Measurement:** responsible for allowing the machine operator to place calliper marks on the displayed image and then calculating the distances between them in order to measure features, calculate volumes, etc. It is also

responsible for other types of measurements obtained from the front-end electronics, such as Doppler flow.

- **Beamformer Control:** management of the front-end electronics to allow user control of the beamforming circuits. This program translates the main control settings into values that are communicated to processors in the front-end electronic via proprietary interfaces.
- **Workflow:** guides the Sonographer through the process of performing an examination by prompting her through a sequence of actions, such as:
 - Entering patient details from keyboard, or looking them up from Hospital Information System (HIS)
 - Selecting preconfigured settings appropriate for the type scan and loading them into the front-end electronics
 - Probe activation and adjusting controls on the user interface to optimise the image by changing the probe and front-end electronics settings.
 - Acquisition of the required images to form clinical judgement and entering comments as well as associated measurements (Doppler flow) into the examination report
 - Conversion images from the front-end cine memory to files that can be stored by the operating system on the hard-disk.
 - Making measurements of features in the images (i.e. ovary dimensions) and providing other annotations (i.e. left or right ovary)
 - Completing a structured report for the examination and storing it in the HIS. This might involve reviewing images previously saved as DICOM files

The back-end computer also allows selection of applications for diagnostics, image review from DICOM files and so forth.

Display

The main display, though usually quite small, is rendered on a high quality colour monitor of the type commonly used with Personal Computers. Very large monitors are seldom used as the ultrasound image is relatively small so during examination they offer no advantage. For example, the images gathered by UKCTOCS are 640

pixels wide and 480 pixels high so can easily be displayed on the 15" monitor provided by the Medison Accuvix XQ machine.

The image is principally grey-scale with received ultrasound echo intensity represented by 8 bits per pixel such that the values of 0 and 255 are displayed as black and white respectively. However, image is actually extended to 24 bits per pixel encoding because it may need to show the red and blue Doppler shift associated with blood flow as an overlay on the grey-scale image. This allows 8 bits per pixel for each of the primary colours (red, green and blue) with areas of grey-scale having the same value in each 8 bit byte – i.e. mid-tone grey is represented by the values Red=128, Green=128 and Blue=128.

The circuitry needed to produce the display monitor's signal can be provided by a standard PC graphics card, though special device drivers may be needed to map the display memory of the card onto the display memory of the beam former in the front-end electronics; see Figure 2-2. Depending upon the probe type, the resulting ultrasound image is displayed either as a rectangle or a sector (pie slice). The probes used by UKCTOCS produce a sector scan and show a cross-sectional view of the uterine Adnexa; see Figure 2-7. The organs closest to the vaginal probe are at the top of the image and those furthest away are at the bottom. Organs on the left side of the image are detected at points on one side of the image plane, and organs on the right side are detected at points on its other end. Rotating the probe simply moves the image plane giving a different view through a section of the Adnexa.

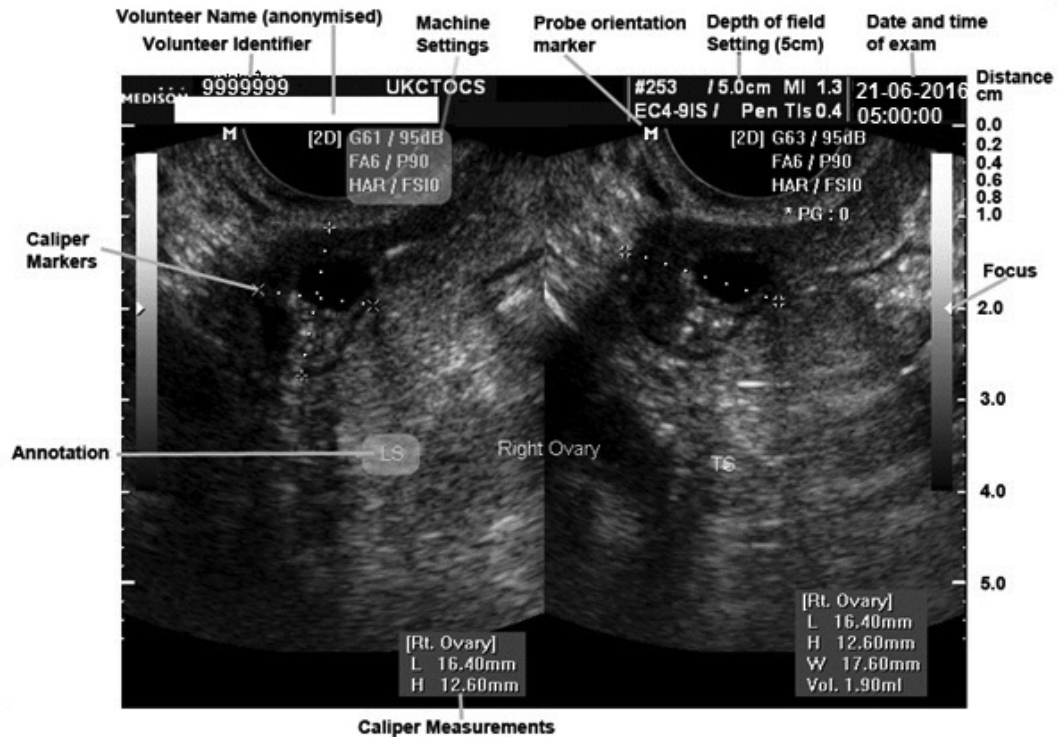


Figure 2-7: Static Ultrasound Image Captured during a TVS Examination with annotations showing the various types of information provided by the ultrasound machine to the sonographer. The calliper marks used by the sonographer to measure the ovary are also shown together with the annotations she provided to identify the image as being obtained from the right ovary.

User Controls for Specific Ultrasound Functions

The sonographer is able to control various other aspects of image creation using settings on the machine, the most important of these are recorded on the static images gathered during an examination, as shown in Figure 2-7, and are given as:

- **Depth of field** – the penetration distance of the image plane as displayed on the monitor. The scale on the sides of the image gives the absolute distance, so that in Figure 2-7 the image plane extends 6cm from the tip of the transducer. Therefore, only echoes whose delay time is less than that corresponding to a 6cm penetration depth are shown. The depth of field should be adjusted so that the ovary appears a reasonable size in the image allowing accurate calliper measurements to be made
- **Focus** – the ultrasound beam should be focused in order to maximise the resolution of the display at the depth of the ovary

- **Power** – the power applied to the transducer determines the amplitude of the transmitted ultrasound waves. This in turn determines the relative amplitude of all the reflected waves allowing the Sonographer to make the display appear brighter or dimmer. The values for Thermal Index (TI) and Mechanic Index (MI) correspond to current power settings and are displayed on the monitor.
- **Dynamic range** – reflected ultrasound waves received by the transducer are converted into a signal that determines the brightness of the corresponding pixels in the display. Therefore, when the Sonographer adjusts the dynamic range this changes the range of pixel brightness in the display giving it more or less contrast.
- **Gain** – the signal generated by the transducer in response to receiving reflected ultrasound waves needs amplification. If the amplification is set is too high then large signals will be clipped. If the amplification is set is too low then small signals will be lost.
- **TGC** – time gain compensation corrects for the attenuation signals suffer as a consequence of coming from reflections deeper in the body; see Front-end Electronics, TGC circuit. This avoids the picture becoming darker as it progresses down the y axis simply because the amplitude of the signal is decreasing with depth.
- **Zoom** – the number of pixels in the display used to represent a given area of tissues in the body can be adjusted in order to increase the size of the object being examined. The zoom setting controls the number pixels between the large scale marks at the edge of the image typically set at 1cm intervals for TVS scanning.

The sonographer adjusts the controls to optimise the image shown on the main display. Typically, as described by Grichnik⁴⁸, this is achieved by setting the depth so the area of interest is centred at a line approximately three quarters of the way down the display and then setting the focus so the centre of this area has the best lateral spatial resolution (if necessary the frame rate is also reduced to create a number of focal zones within this area – i.e. reduction of temporal resolution). Next, the TGC is set to obtain a smooth grey-scale picture and finally the sonographer

adjusts the zoom, gain and dynamic range to respectively magnify the point of interest, optimise detail and enhance contrast.

The user interface of the machine is usually comprised of the main display, keyboard, tracker ball (mouse) as well as various custom controls for specific ultrasound functions; see Figure 2-8. Each input component of the user interface has its own device driver software which allows operator inputs to be fed to applications under the overall control of the back-end computer's operating system. Similarly, application outputs are fed by the operating system to displays, indicator lamps, speakers and so forth.



Figure 2-8: Medison Accuvix XQ User Interface is designed to facilitate operation of the machine in a clinical setting. The middle of the picture shows a touch panel containing controls that the machine provides according to the stage of the examination in order to reduce the number of static controls that the sonographer needs to operate. © Samsung Medison Co., LTD. Reproduced with permission.

In recent years considerable importance has been given by manufacturers to the ergonomics of their user interfaces in recognition of the fact that superb imaging features are of little use if the operator cannot understand how to use them, or finds the setup too complex to perform during a 15 minute TVS examination.

Consequently, most manufacturers complement the slider and rotary knobs of their

machines' standard controls with touch-screen units containing virtual controls that can be configured to help the operator through complex tasks. Indeed, manufactures like Carestream⁴⁹ have replaced all the controls, including mouse and keyboard, with a large touch-screen.

System Integration

Most large NHS hospitals have implemented a central archiving facility for digital images acquired from their radiology departments. These systems are termed Picture Archiving and Communication Systems (PACS) and implement various standards for the viewing, storing, retrieval, management and communication of medical image data and associated examination records. For example, images are stored in an international standard format called Digital Imaging and Communication (DICOM) which managed by the National Electronic Manufacture's Association, a US trade body. Details of the DICOM file structure are given in Appendix H.

Images can be transmitted from an ultrasound machine to the PACS using a standard network protocol with integration provided at the application layer using a standard software interface. The ultrasound machine may also be integrated with the Hospital Information System to access patient record data using standard messaging services such as HL7. The back-end computer will typically support these various standards using commercial device drivers to implement the necessary communication stacks, as well as property software libraries integrated with custom software applications providing the required functionality. However, UKCTOCS did not implement a PACS, so system integration was not an important consideration in respect of their machine selection.

Selection of Ultrasound Equipment

Manufacturers do not generally publish detailed specifications relating to the performance of their ultrasound machines. In part this reflects the difficulty of providing meaningful values for parameters like machine resolution (see Resolution Integral). It also reflects the importance of proprietary algorithms implemented in the signal processing pipeline which may serve to enhance the image by removing speckle or provide super resolution by combining frame data in the cine memory.

Therefore UKCTOCS did not specify machine performance in quantitative terms. Instead machines were selected using selective criteria such as qualitative image quality, ergonomics, feature availability and so forth^{50,33}.

Machines used by UKCTOCS

The majority of trial TVS scans were performed using dedicated ultrasound machines belonging to UKCTOCS. Two types of machines were used. Between 2001 and 2007 the Kretz SA2000 machine (Kretztechnik AG, Zipf, Austria) was used for scanning with a phased array transvaginal probe (4-9 Mhz). However, new machines were procured in 2007 so the Medison Accuvix XQ machine (Samsung Medison, Seoul, South Korea) with a transvaginal probe type EC4-9IS (4-9 MHz) was used for scanning performed after 1st January 2008. The Accuvix XQ machine benefited from improved imaging capability and was used until the end of the trial in 2011. It is shown in Figure 2-9.



Figure 2-9: Medison Accuvix XQ ultrasound machine used by UKCTOCS after 1st January 2008. © Samsung Medison Co., LTD. Reproduced with permission.

The Technical Reference manual for the Medison Accuvix XQ is mainly a list of bullet point features without further detailed description. For example, calliper measurement in B-mode lists the types of measurement the machine can produce (distance, angle, area, ellipse, circumference, volume), but does not give details of

their range, accuracy or precision. Attempts to obtain further information from the manufacturer were unsuccessful, but it was suggested informally that precision for UKCTOCS measurements were unlikely to be better than ± 1 mm, even though values were displayed to 0.01mm (see Figure 2-7). Inquiries made at a trade show (BMUS, 2015) also failed to yield any specific information about specific technology employed in machines, though ergonomics were a clear differentiating factor between the machines exhibited.

The decision by UKCTOCS to replace all its ultrasound machines with the Accuvix XQ model was made on the basis of a subjective assessment that the new machines would deliver higher quality images. Although anecdotal comments by sonographers and the experts supervising them suggests the Accuvix XQ did indeed result in a substantial improvement in image quality, the lack of any systematic study comparing the quality of images in the URA before and after 2008 means that the benefits have not been quantified. Furthermore, examination of ovary visualization data in the paper by Sharma et al³⁰ indicates that the new machine may have actually caused a reduction in exam quality in the first year of its use, though the trend reversed in subsequent years (see discussion in Chapter 3). This suggests that some sonographers initially found it difficult to perform TVS exams with the same levels of skill when using an unfamiliar machine, so supporting the case that factors which help sonographers make better use of the machine may also contribute to quality of TVS scanning, at least in terms of ovary visualization.

Development of Ultrasound Technology and Future Trends

The use of ultrasound in obstetrics and gynaecology can be traced to the seminal paper by Donald et al in 1958⁵¹ and within a decade companies like Kretztechnik AG (now Samsung Medison) were producing relatively sophisticated B-mode machines. By the late 70's multi-element linear and phased array scanners had largely superseded mechanical real time scanners for transabdominal use, though the first commercial endovaginal probe released by Kretztechnik in 1985 still had a mechanical design.

The semiconductor revolution during the 70s and 80s provided the technology that made possible the sophisticated signal processing needed for beam forming and

multi-channel arrays. However, the most significant improvements in image quality came in the 1990s, when the availability of sophisticated microprocessors coincided with significant improvements in software capability as a result of emergent new programming languages like C++. This allowed most of the analogue circuitry in probe and front-end electronics to be replaced by digital electronics, delivering a step-change in performance as well as a wide range of new imaging features. In comparison development in sequent years was more incremental, though never-the-less significant in some areas. Therefore it can be expected that future technical improvements in ultrasonography will continue to be incremental unless the industry is impacted by new disruptive technology as it was in the 1990s.

Undoubtedly further developments will continue to occur in probe and cable design leading to better dynamic range and improved signal to noise ratios allowing future machines to process signals with enhanced contrast and sensitivity. Further digitalisation of the signal pathway in the probe will also help in this regard as well as permitting use of the thinner and more robust cables associated with digital data transmission. However, the area with the most obvious potential for progress is the front-end electronics whose signal processing pipeline seems likely to benefit from the introduction of more powerful GPUs and DSPs as well as the implementation of new and better algorithms for processing the data from the probe. This will encourage the development of technologies like histoscanning with the potential to present additional information to the sonographer during TVS scanning, such as whether a lesion detected in the ovary is likely to be malignant. It is also possible that information about the quality control of the examination will be provided in real-time, for example further development of the work presented in Chapter 9 may allow assessment by the machine of the probability that both ovaries have been visualised. However, the utility of such features is clearly dependent on the sonographer being able to use the equipment efficiently. In this respect, factors such as ergonomics and back-end computer support for scanning workflow should help improve sonographer efficiency as suggested by Sanyal et al⁵². It remains to be shown what impact the implementation of a standardized workflow for TVS scanning might have on its quality control, but there is obvious potential for improvement in this area.

Summary

It is accepted that the ability of the sonographer is a major factor that determines the quality of a TVS examination. However, there is also an assumption that suitable equipment will be used during scanning. In this regard it is important to understand the nature of ultrasound and the role of the various components in the signal pathway from probe to display.

Further development of image processing in the front-end electronics as well the implementation of software to support the sonographer in the back-end computer have particular potential for helping to improve the quality of future TVS examinations. However, it is likely that the selection of ultrasound machines for future large-scale screening programmes involving TVS will continue to be dominated by the same type of subjective criteria that was used by UKCTOCS. It is possible that more objective measures such as the Resolution Integral will allow new machines to be selected on the basis of the compliance of their specifications with common standards. In addition, factors besides imaging performance may influence machine selection due to better understanding of their importance to TVS examination quality. Ergonomics, workflow support and real-time quality control have particular relevance in this respect.

Ultimately, the quality of an ultrasound machine needs to be judged in terms of examination outcome. All other things being equal, would the same sonographer detect an abnormality in an ovary using one machine, but not in another? Similarly, does the machine allow sonographers to assess an ovary as normal, that otherwise might be considered boarder-line abnormal? It is for this reason that objective quality control of the images is important as it provides evidence of both sonographer and machine performance as discussed in the next chapter

Chapter 3: Managing the Quality of TVS Scanning

Introduction

This chapter starts by tracing the development Quality Improvement (QI) from its origins as simple control charts for allowing the statistical monitoring of a manufacturing process for telephones to the more recent emergence of Six-Sigma methodology which has been successfully applied in multiple industry sectors to deliver high quality defect free products and services. The chapter then explores existing work in implementing QI within healthcare and identifies potential sources of support and expertise for QI in respect of TVS scanning. It concludes with a review of current quality practices in medical ultrasound in respect of people, equipment and procedures – particularly those related to using TVS as a screen for ovarian cancer. In this way a context is provided for the application of the tools described in the following chapters in terms of their use in a QI programme.

Concepts of Quality Assurance, Quality Control and Quality Improvement

Quality Assurance (QA) is a process. It can be understood in terms of a cycle that starts by setting standards, continues by measuring practice, compares practice against standards, identifies the required changes, implements the changes and then re-evaluates practice before finally repeating the cycle by setting or adjusting standards again. Quality Control (QC) is concerned with measuring practice within this QA cycle so subsequent parts of the cycle can drive improvement on the basis of reliable metrics that reflect the actual quality of a product or service.

Quality Improvement (QI) has a similar iterative approach to QA, but one driven by evidence-based benchmarks, experimentation and objective measurement. In QI the impact of change is assessed with aim of delivering continuous improvement against specific objectives. This approach contrasts with QA which simply serves to ensure quality standards are met without any implicit need to deliver goals like improved care or reduced costs.

In the context of TVS scanning, QA might set a standard for sonographer performance in terms of exceeding a certain score for a given metric like visualization rate (VR) – the proportion of scans in which both ovaries were seen by the sonographer. QC would then measure this metric for each sonographer to assure quality of their work conforms to the stated standard, say visualising both ovaries in 60% of exams. In contrast QI might set a goal of detecting ovarian abnormalities during the first TVS scan for 99.99% of the women who receive treatment for ovarian cancer; this might be termed First Scan Detection Rate (FSDR). The baseline FSDR would then be measured and experiments conducted to find objective QC metrics that might improve FSDR. The sonographers would receive help to improve their scores for such metrics, with new metrics being introduced periodically to reach the next level of FSDR improvement. In this way QI for TVS scanning would progressively reduce the number of women diagnosed with ovarian cancer who did not have ovary abnormalities detected during their first scan and so suffered delay in their treatment.

Clearly, different values might be set and different goals specified when attempting QI for TVS in clinical use. However, the fundamental approach would be the same in terms of incrementally improving the process by iterating through a succession of initiatives in order to satisfy an overarching goal which (unlike VR) has tangible value to patients or the organisation. Each of these initiatives would involve the setting of QC standards which could be measured by objective metrics with the aim of driving improvement towards the next level of quality.

Use of Statistics in Quality Improvement – a historical perspective

The application of statistics to the problem of quality improvement (QI) was first suggested in 1924 by W. A. Shewhart of Bell Telephone Laboratories. He sought to improve the reliability of his company's transmission systems by reducing variation in the manufacturing process, but realised that only part of the variation was caused by natural (random) causes such as might arise from a machine making components within a range of tolerances. The other part of the variation arose from changes that were occurring within the process due to effects like machine wear. He termed such effects 'special' variation. Therefore to help decide when it would be beneficial to make adjustments to the manufacturing process, he proposed creating a control chart for variables; see Figure 3-1. This was constructed by measuring a variable associated with the quality characteristics of the system and then calculating the mean and standard error so control limits could be set at ± 3 standard errors from the mean. In this way, the process would be adjusted only when the values of the variable exceeded the control limits or satisfied some other criteria indicating a non-random cause. Shewhart's work was extended by others including W. Edward Deming to provide the foundation for modern Statistical Quality Control (SQC) methods which are now incorporated in the various approaches to QI adopted by many industries throughout the world.

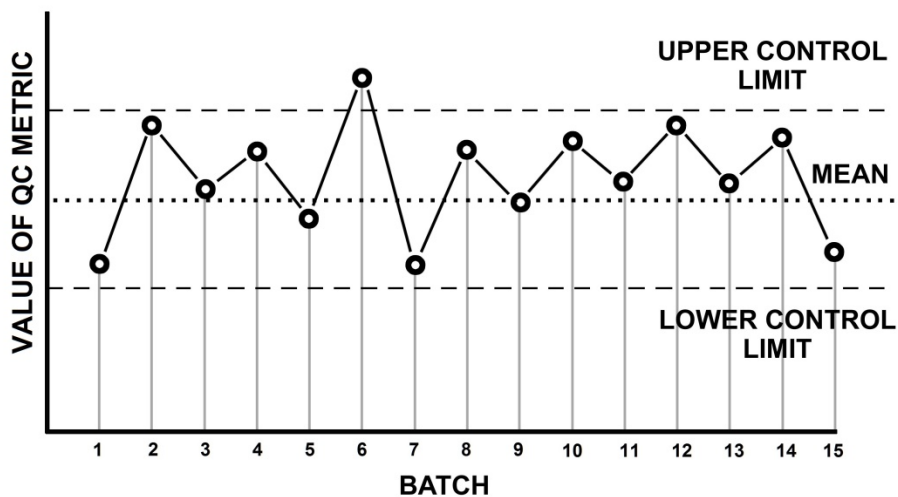


Figure 3:1 Control chart for variables. The values of a Quality Control (QC) metric vary by batch number, but it can be seen that for most batches the values are within the upper and lower control limits which correspond to ± 3 standard errors from the mean. However, the value of the QC metric for batch 6 lies outside this range suggesting a source of non-random variation in the manufacturing process. The values for subsequent batches suggest that problem was addressed.

The advent of effective quality control tools caused significant interest in the challenges of improving quality within organisations. This was particularly the case in Japan as new ideas about quality management had a transformational effect on the growth of companies such as Nissan, Nippon and Toyota. Building on the work of Deming, Kaoru Ishikawa developed the concept of Quality Control circles at Nippon to bring together small groups of workers in the same field in order to identify, analyze and solve problems associated with their work. This eventually led to the adoption of Total Quality Management (TQM) as a way of increasing quality through the systematic analysis and improvement of work processes. Quality became understood as something that required improvement and was the responsibility of everyone in the organization from the CEO to the production line worker.

TQM is primarily a cultural approach to enhancing quality, though it may employ SQC tools and techniques to provide the data needed to drive improvement. An organization may also define its process for improving quality in terms of a standard such as ISO 9000 which allows it to obtain external validation and certification. However, this in itself does not result in an organization reducing defects and errors

in their products and services to levels at which they are essentially eliminated. It simply defines the process whereby such goals may be achieved.

Quality Improvement using Six-Sigma Methodology

Six-Sigma is a methodology originally developed by Bill Smith at Motorola in 1986. It builds on Deming's work, particularly his Plan-Do-Check-Act (PDCA) cycle⁵³ which describes an iterative cycle of making a plan for improvement, implementing the plan, looking at its results and then using this information to improve the process. Six-Sigma is now commonly used in a number of industries including healthcare⁵⁴. It is concerned with the delivery of a high quality and defect free service or process and describes phases of activity to achieve specific value targets. These phases of activity are led by a quality management team using statistical methods and tools to

- a) reduce relevant discrete variable (attribute) outcomes to a zero defect level
- b) ensure all measurable continuous variable outcomes fall within upper and lower tolerance limits set at plus or minus six standard deviations from their mean value.

Measuring Process Capability

Measurable outcomes may be readily applied in the case of a process to manufacture a part like a piston ring for a diesel engine. In such a case the key requirement might be that its diameter should be 50.0 mm within a specified tolerance of ± 0.1 mm as measured with a standard measurement device, like a Vernier calliper. If the process is predictable and stable then the fabrication errors in respect of this dimension will be normally distributed and centred about a mean of zero error. Therefore the mean diameter of all piston rings made by the process would be 50.0 mm and the frequencies of parts with other diameters would follow a normal distribution curve such that 99.73% of them would be represented by area lying below the curve within ± 3 standard deviations of the mean value. Such a process could be said to be operating at a three-sigma level; see the left curve in Figure 3.2. That is to say 99.73% of parts would be fabricated within the specified tolerance of ± 0.1 mm.

Further improvement might then be made to the process such that the number of parts having a diameter within the specified tolerance would be represented by the area lying under the normal distribution curve within ± 6 standard deviations of the

mean value. The process would then be said to operate at a six-sigma level; see the right curve in Figure 3.2. This means 99.99966% of piston rings would be manufactured within the specified tolerance of ± 0.1 mm which corresponds to only 3.4 parts in every million being substandard. Data about the number of substandard parts produced by a process is typically expressed in terms of defective parts per million opportunities (DPMO), so a process operating at six-sigma has a DPMO of 3.4. Similarly data gathered about the actual DPMO for a process that is predictable and stable allows measurement of its sigma level; a metric which can be used to drive improvement.

Removal of Special Variation

The normal distribution curves shown in Figure 3.2 reflect naturally occurring variation in the piston ring manufacturing process, such as might result from randomness in the ability of the milling machine to achieve the required accuracy when cutting the part. However, the failure of a particular operator to setup the machine properly might cause a significant number of piston rings to be made with diameters that exceeded the specified tolerance on the days he worked in the factory. This creates a source of special variation as the DPMO is influenced by the presence or absence of the rogue operator. Therefore any attempt to measure the process in terms of sigma level would be futile because the standard deviation calculations assume normally distributed data. Only after the identification and elimination of all special sources of variation would the DPMO data become normally distributed – i.e. all defects caused by naturally occurring events like as the innate randomness of the milling machine cutting tolerance.

When a process has only naturally occurring variation it becomes predictable and stable which is a precondition for measuring its sigma level. Therefore any quality project that uses sigma level as a metric to drive improvement must first remove all special sources of variation in the process so the source variation in the data gathered is entirely natural. Only then should improvement be attempted such as buying a new machine with higher cutting accuracy that can make parts with lower natural variation in tolerance.

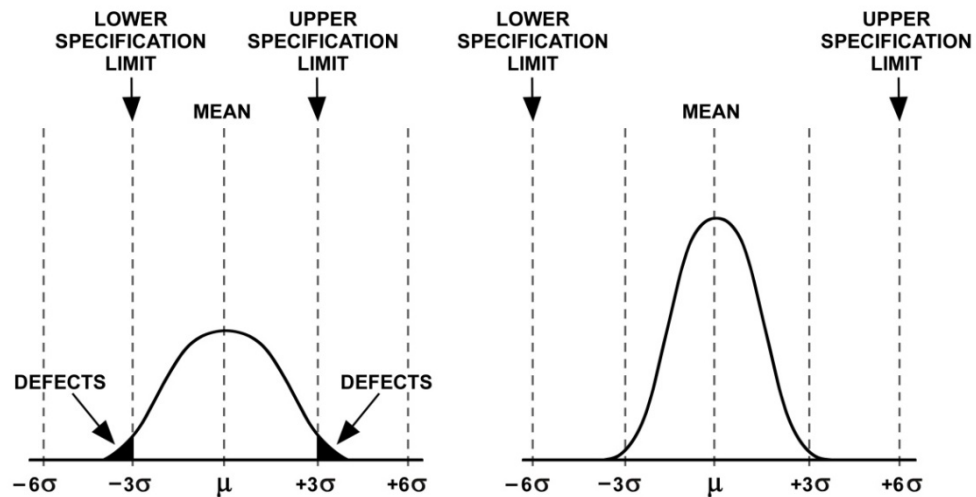


Figure 3:2 Six-Sigma – Zero Defects. In a process that has eradicated the source of all non-standard variation only random variation remains. Therefore the frequency of parts having a quality control metric which varies from the mean will have a normal distribution. The left chart shows the output of such a process with specification limits at \pm three standard deviations from the mean which produces 2.7 defective parts for every thousand made (0.27%). However, the right chart reflects improvements to this process with specification limits at \pm six standard deviations from the mean which produces only 3.4 defects for every million parts made – i.e. a process which generates almost zero defects.

Methodology - DMAIC

The advance made by Six-Sigma lies not just in the setting of higher targets for quality, but also in the formalisation of a methodology for implementing continuous improvement of quality in an organisation. When a Six-Sigma project is setup to improve an existing process^a the methodology requires successive iteration through the following series of phases:

- **Define:** the process is specified as a system and the people who use it (customers) are identified. The requirements of the customers are also stated so the process can be refined to satisfy these needs in a way that gives

^a Design for Six-Sigma is used for a new process to define, measure, analyze, design and verify (DMADV).

measurable improvements in process quality in respect of the sought-after project goals (value targets).

- **Measure:** metrics associated with the desired improvements in process quality are gathered. If normally distributed they are used to calculate the current process capability.
- **Analyze:** the collected metrics are inspected to explore their relationship with quality in order to suggest changes in process that may result in its improvement
- **Improve:** the process is altered to optimise quality as suggested by the work in the analysis phase. Typically, this involves the elimination of variation.
- **Control:** monitor the process to confirm that the changes have resulted in the anticipated improvements in quality whilst still satisfying customer requirements.

During each of the Six-Sigma project iterations the knowledge gained is used to perfect an understanding of the customers and their requirements as well as the specification of the process and the setting of project goals. Adjustments are also made to the process and the resulting changes in its capability are monitored so they can be used to inform future work. In this way quality is incrementally improved until the cost of further quality improvement cannot be justified. At this point the project is terminated so its resources can be employed in other areas with the potential for a better return on investment. However, monitoring of the key metrics of the process will continue so any change in quality can be identified and corrected, if necessary by creating a new Six-sigma project.

Tool Selection

Established management tools are used in the various phases of a Six-Sigma project. For example, during the ‘measurement’ phase SQC control charts are commonly used to monitor variables that might cause special variation. In the ‘analysis’ phase this data will be tested by statistical methods like Chi-Squared to prove hypotheses about causes of variation so key sources can then be identified using tools such as root cause analysis⁵⁵. In this way suitable changes can be made during the ‘Improve’ phase and their impact observed during the ‘Control’ phase. Clearly, well-constructed experiments are necessary for teams to produce meaningful results and

avoid trap such as the impact of observation; Hawthorn effect⁵⁶. Therefore Six-Sigma projects often use techniques in their work generically referred to as Design of Experiments (DoI) with derivations proposed by Taguchi⁵⁷ and Shainin⁵⁸.

In addition to statistical tools the ‘improve’ or ‘control’ phases may use tools from Lean Manufacturing⁵⁹ such as Kanban cards⁶⁰ (signals for stock replenishment) or Andon displays⁶¹ (process status communication). Indeed, the term Lean Six-Sigma⁶² is used to describe a project that uses such tools to implement the lean ideals of removing waste (muda⁶³) within a Six-Sigma framework.

One of the key challenges when implementing a Six-Sigma project is finding the right tools to help achieve its goals. Each project will differ in this respect so requires team members with experience in using a wide range of management tools and statistical methods. Knowledge is not a substitute for experience in this regard and for this reason any successful Six-Sigma project will value its people and their interactions over the prescriptive use of process and tools.

Accreditation and Training

Six-Sigma is underpinned by rigorous training in the theoretical and practical application of SQC and associated problem solving tools delivered by accredited experts who have a proven record of implementing Six-Sigma change programmes.

The lowest level of accreditation is a ‘green belt’ – i.e. someone who has completed the necessary training and has worked in a Six-Sigma team. The next level is a ‘black belt’ which is awarded to someone who has led a process improvement team. The higher level of ‘master black belt’ is given to someone who has coordinated Six-Sigma programmes across an organisation and is qualified to train ‘green belts’ and ‘black belts’. These incremental levels of training reflect the importance of viewing quality as a cultural issue for an organisation rather than simply as the introduction of various tools and targets. The aim is to achieve defect free services and processes by transforming the entire organisation into one that fully embraces Six-Sigma. Unfortunately not all organisations are able to implement such a significant cultural shift.

Implementation Challenges and Criticisms

The successful implementation of Six-Sigma projects depends on obtaining support from the CEO and other people in the organisation's upper management team. These senior executives provide the vision, resources and delegated authority necessary for their Six-Sigma teams to succeed. They will also appoint individuals from amongst their ranks to act as champions for such teams, i.e. someone who can provide strategic advice to the black belt as well as act as an advocate for the project at a high level in the organisation. All this implies that the chief decision makers have fully embraced Six-Sigma and are prepared to make the necessary alterations to their culture. Providing the justification for making such changes is typically a key challenge.

Evidence for the benefits will need to be quantified in terms of cost savings that would result from quality improvement as well as less tangible gains such as reputational enhancement, better staff moral and delivering an improved service or product to their customers. Such benefits will then need to be balanced against the cost of implementing Six-Sigma which may be substantial. The quality improvement teams will need to be assembled, either by moving existing staff from other work or by recruiting additional head-count. Each team will also need to recruit a black-belt having the skills and experience necessary to lead it. In addition there are training costs to consider as well as the costs of employing external consultants to provide oversight and monitoring of the work. Six-Sigma costs are not just related to the setup and running of Quality Improvement (QI) teams, for the organisation will also need to fund the expenditure associated with changing processes identified as having scope for improvement. This may mean setting-up a separate 'change' project with its own budget for staff as well as buying new equipment, training operators, or providing other types of resources.

One of the main criticisms of Six-Sigma is a tendency for teams to focus on the process and tools of quality improvement rather than its outcomes. This is a particular issue when Six-Sigma is being introduced into an organisation as people in the teams are learning new skills and vocabulary as well as undergoing a change in culture that may overwhelm them for a time. This may result in the team following the steps prescribed by the methodology without giving too much thought about

what they are seeking to achieve. For example, some rare events are not normally distributed so their occurrence cannot be predicted by traditional statistics, but this does not prevent some Six-Sigma teams collecting and analysing the associated metrics. The team when lead by an effective black-belt should avoid such a pitfall, but as Taleb⁶⁴ observes the blind use of statistical tools and methods, even by experts, is common-place.

Approaches to Quality Improvement in Healthcare

There is significant interest in the potential of quality improvement (QI) to allow healthcare in the UK and elsewhere to deliver better patient care at lower cost⁶⁵. However, the methods adopted to improve quality in healthcare are diverse.

Quality Improvement Studies

A systematic review of the application of QI methods from manufacturing industries in surgical healthcare was conducted by Nicolay⁶⁶ et al in 2011. The authors searched a number of databases and found 34 articles about studies that satisfied its inclusion criteria. These studies took a number of different approaches to QI including Six-Sigma (5), Lean Six-Sigma (1), Lean (4), TQM (5), PDCA (5) and SQC (5). In addition 9 studies described continuous quality improvement (CQI) initiatives which Hughes et al⁶⁷ describe as being the much the same as TQM, but applied in a healthcare setting. This systematic review concluded that “*QI methodologies from industry can have significant effects on improving surgical care, from reducing infection rates to increasing operating room efficiency*”. However, it was noted that the studies lacked compelling evidence such as might be generated by randomised multicentre studies.

An earlier systematic review was conducted by Thor⁶⁸ et al in 2008. It involved 311 studies in healthcare published between 1990 and 2004 concerned with Statistical Process Control (SPC) – i.e. processes for improving quality that use control charts as a core tool as well as continuous improvement and DoI. It did not report on whether SPC was used in conjunction with other methodologies such as Six-Sigma and acknowledged that the rigour of the methodology applied in many of the studies could not be determined. However, it did find that there were reports of “*substantial*

benefits of SPC application” and concluded that SPC is a “*versatile tool which can help stakeholders to manage change in healthcare and improve patient’s health*”. It noted that “*the key [to widespread adoption] then is to develop or recruit the expertise necessary to use SPC correctly and fully and to make SPC easy for non-experts to use*”. It also noted that the majority of SPC studies had been conducted in the USA which was considered to reflect the relative adoption of quality improvement in healthcare globally.

Davidoff et al⁶⁹ report on a refinement of earlier work to define Standards for QUality Improvement Reporting Excellence (SQUIRE). The introduction states “*A great deal of meaningful and effective work is now done in clinical settings to improve the quality and safety of care. Unfortunately, relatively little of that work is reported in the biomedical literature, and much of what is published could be more effectively presented*”. The proposed standards aim to provide a framework for “*reporting formal, planned studies designed to assess the nature and effectiveness of interventions to improve the quality and safety of care*”. However, a search of Web of Science (1/3/16) using the topics ‘SQUIRE’, ‘Quality’ and ‘Healthcare’ returned 22 articles of which only 3 articles used SQUIRE as a framework for presenting original research.

Kaplan et al⁷⁰ report on a Model for Understanding Success in Quality (MUSIQ) to provide a theory of context for healthcare QI. They note that the benefits implementing QI in healthcare has been variable with some projects delivering significant improvements in processes or patient outcome, whilst other produced only modest gains. They suggest this mixed success is due to “*the effects of context on the successful application of QI methods, not the efficacy of the methods themselves*”. Context in this respect is defined as “*characteristics of the organisational setting, the environment, the individual, and their role in the organisation or QI project*”. The model they propose allows contextual factors such as motivation, leadership, and culture to be considered separately to the actual methodology adopted or the nature of the clinical interventions. The model also considers the way such factors operate at multiple levels within an organisation and the impact of external influences.

MUSIQ identifies 25 key contextual factors which are considered to determine QI project success. These factors are defined and can be measured so that their individual influence on outcome can be modelled using data from multiple projects both within the same organisation and across different healthcare organisations. In this way it is hoped that QI teams will develop a better understanding of the influence context has on their projects separately to the impact of the methodology or clinical treatment. The model should also provide guidance to people at all levels in healthcare about the steps they can take to optimise the results of specific QI activities in terms of addressing context issues.

Quality Improvement Guidelines for UK Healthcare

A Google search was made on 1st March 2016 for internet links containing information about quality improvement initiatives in healthcare; search term “Quality Improvement Healthcare UK”. It yielded more than 14 million results so only the first fifty were considered as they were ordered in terms of importance. Results were included for review if they satisfied the criterion of containing detailed information about an approach to quality improvement or quality assurance in UK healthcare. Results relating to courses or conferences about QI were excluded from review. The primary aim of the review was to identify the key drivers of QI in the UK. Its secondary aim was to find the main organisations providing QI guidance.

Key Drivers of QI in UK

The gap between NHS income and expenditure is estimated at £2bn in the financial year 2020/21 rising to £9bn in 2030/31⁷¹. This creates considerable pressure for the NHS to reduce its costs. At the same time the NHS is under considerable pressure to deliver better patient care, particularly as a consequence of problems uncovered in the last decade leading to enquiries conducted at the Mid Staffordshire Trust⁷², Morecambe Bay⁷³ and elsewhere.

Quality improvement programmes at companies like Toyota⁷⁴ and Boeing⁷⁵ have long been credited with deliver better products at lower cost. Translating this success into delivering better patient care at lower cost is a key driver for the implementation of quality improvement in the NHS⁷⁶. This is because QI is one of the few options

the NHS has for making a step-change in the way it operates to secure its long term survival.

Key Organisations Proving QI Guidance in UK

Provision of NHS services in the UK is divided geographically into NHS England, NHS Scotland, NHS Wales and the NHS in Northern Ireland. The Scotland Act (1998) devolved certain powers from the English Parliament including responsibility for NHS service in Scotland. A similar devolution of powers has been enacted for Wales (1999) and Northern Ireland (2009). Therefore each country in the UK has a separate and largely autonomous body with statutory responsibility for providing healthcare to its citizens. The websites associated with each of these bodies describe their different approaches to QI:

- **NHS Scotland** formed the ‘Quality Improvement Hub’⁷⁷ in 2010. It is a national collaboration among special health boards and Scottish Government Health Directorates. Amongst the resources it provides is an introduction to Statistical Quality Control (SQC) and Plan-Do-Study-Act cycles (PDSA).
- **NHS Wales** created ‘1000 Lives Improvement’⁷⁸ as its nation improvement service. This site delivers the national learning programme (Improving Quality Together) to provide a “common and consistent approach to improving quality of services in NHS organizations across Wales”. The programme provides a Bronze, Silver, Gold and Board level accreditation (similar to Six-Sigma belts) with a model for improvement is based on PDSA methodology.
- **NHS in Northern Ireland** delegates their quality improvement activities to the ‘Regulation and Quality Improvement Authority’⁷⁹ (RQIA) which was formed in 2003. The RQIA contributes to Quality 2020⁸⁰ initiative; a 10 year strategy started in 2011 to protect and improve Health and Social Care (HSC) quality in Northern Ireland. The Quality 2020 Implementation plan included an Improvement Methods Project (1A) to ‘identify, evaluate and promote available quality improvement methodologies for application across all HSC services’. However, the results of this project cannot be found on the HSC website⁸¹.

- **NHS England** created the ‘Sustainable Improvement Team’⁸² in November 2015 to replace the ‘NHS Improving Quality’⁸³ organization it formed in 2013 after closing down the ‘NHS Institute for Innovation and Improvement’⁸⁴ established in 2005. The QI approach of NHS England has evolved since 2005 and is now based on a team of experts working to embed improvement across the health and care system to support NHS England’s priorities. They refer to the old ‘NHS Improving Quality’ website for specific information about how that might be achieved, but no reference to any particular methodologies was found.

Provision of QI Guidance in England

The UK Social Care Act 2012⁸⁵ made the most fundamental changes to the way the NHS in England was organised and run since it was created in 1949. It created the Care Quality Commission (CQC) as the regulator of health and social care services in England. It also created NICE (National Institute for Health and Care Excellence) as a body for “*giving of advice or guidance, provision of information or making of recommendations about any matter concerning or connected with the provision of— (a) NHS services, (b) public health services, or (c) social care in England*”. In addition the Act created the Commissioning Board (NHS England) and included in its mandate a requirement to “*exercise its functions with a view to securing continuous improvement in the quality of services*” it also states “*In discharging its duty under subsection (1), the Board must have regard to the quality standards prepared by NICE under section 234 (Part 1, Section 23, paragraph 13E)*”. Therefore NHS England has a statutory duty to implement continuous quality improvement and adhere to standards created by NICE in this regard.

National Institute for Health and Care Excellence (NICE)

NICE doesn’t recommend a particular QI approach, but instead publishes quality standard topics about delivering quality care for a range of particular medical issues. For example, NICE Quality Standard 18⁸⁶ concerns Ovarian Cancer (OC). This topic gives a general introduction to OC and is followed by eight quality statements which range from diagnosis to surgical staging. Every quality statement defines an aspect of care to be delivered, its expected quality and how this quality is measured.

Therefore each topic provides a structure for care delivery including targets for its quality. These targets are based on the best available evidence and are considered aspirational, but achievable. In this sense they can be said to drive quality improvement.

In addition to publishing quality standard topics, NICE also publishes advice about audit and service improvement⁸⁷. This includes spreadsheets for evaluating current practice and planning activity for improvement in the case of 155 particular medical conditions. It also includes a general template for quality standard service improvement which can be used for medical conditions not listed.

No information was found on the NICE website about QI methodologies (April 2016). The only assistance identified for the implementation of improvement was a statement under the heading ‘Models for Change’ in Practice Guide PG1⁸⁸ which referred to the ‘Centre for Excellence and Outcomes in Children and Young People's Services’ website⁸⁹ (now archived) for details about ‘commonly used models of service improvement and the theoretical basis underpinning them’.

National Quality Board (NQB)

The National Quality Board (NQB) was formed in response to recommendation made in the Darzi report⁹⁰ ‘High Quality Care for All’ presented to the English Parliament in June 2008. It has no statutory authority, but is the only forum where bodies with responsibility for quality of care can come together to develop unified strategy. The NQB had its first meeting in March 2009.

In April 2013 The NQB published “Quality in the new health system”⁹¹ describing how the newly reorganised NHS should maintain and improve quality. This report reaffirmed the NQB’s role in providing leadership and system alignment for quality as well as asserting that “*the most effective mechanism for preventing quality failure was for organisations and individuals providing care to be continuously striving for improvement*”. The report also explained its approach to continuous quality improvement in terms of the the Quality Governance Framework (described in Chapter 10).

In October 2014 the NHS England published the 'Five Year Forward View'⁹² which included a commitment to re-energise the NQB. Therefore in 2015 the NQB was re-established with new leadership and membership to make it more clinically and professionally focused. It has subsequently formed the subgroup 'Leadership Development and Improvement Board'. In a paper⁹³ presented to the NQC meeting of 28th October 2015 it was acknowledged a single national strategy for improvement and leadership in NHS England was needed. The report also suggested that leadership in service improvement was an area for development in NHS England. There was no mention of QI methodologies.

Provision of QI Guidance by UK Healthcare Professional Bodies

The UK has a large number of healthcare professional bodies. Within the field of ultrasonography the main professional bodies are Royal College of Obstetricians & Gynecologists (RCOG), Royal College of Radiologists (RCR), Society & College of Radiographers (SCoR) and the Royal College of Midwives (RCM). There are also a number of organisations based in the UK that do not have registration requirements for professional membership, but nevertheless are concerned with promoting professional standards in ultrasonography. These include the British Medical Ultrasound Society (BMUS) and the British Society for Gynaecological Imaging (BSGI).

The Royal College of Radiologists (RCR) website⁹⁴ describes quality improvement in terms of five general objectives targeting patient benefit. Their members should deliver services that are safe (avoid harm), effective (have proven results), person-centred (tailored to needs and values of individuals), timely (given when needed), equitable (based on clinical need without regard for age, ethnicity or postcode). In addition there is the further objective of targeting the broader benefit to society by their members delivering services that are efficient and avoid wasting resources– i.e. delivering the same care for less cost. These six objectives are shared by other bodies for healthcare professionals which are part of the consortium leading the Healthcare Quality Improvement Partnership⁹⁵ (HQIP).

Recognising NICE's redefinition in 2002 of clinical audit as a quality improvement process, HQIP published a guide in 2012 to using QI tools to drive clinical audits⁹⁶.

Its introduction states *'It is unclear if clinical audit committees and staff have understood fully the implications of the repositioning of clinical audit in the domain of quality improvement'*. The guide continues by implying that many healthcare professionals are confusing the sort of Quality Assurance (QA) activities they carried out in the 1990s with the modern approaches to QI being promoted by HQIP. Therefore it makes a clear distinction between the application of QA models and QI models in clinical audit.

Quality Assurance of Medical Ultrasound

The processes used to provide the QA of medical ultrasound can be classified under the three broad headings of sonographer employment, equipment and procedure. In this way a patient can have confidence that a) they will be examined by someone with the appropriate level of training and experience b) using equipment which is suitable, safe and properly maintained and c) following a procedure which is appropriate for delivering the required care.

Quality Assurance of Sonographers

Medical doctors specialising in ultrasonography in the UK are registered with the General Medical Council (GMC) and will have completed training coordinated by the Royal College of Obstetricians & Gynaecologists (RCOG)⁹⁷. Most other people working in the field are registered either with the Health and Care Professions Council (HCPC) as a qualified radiographer or with the National Midwifery Council as a midwife⁹⁸. However, sonography is not recognised as a profession by the HCPC and there is no legal requirement in the UK for someone to hold a recognised ultrasound qualification in order to practice as a sonographer. Therefore responsibility for ensuring the quality of sonographers ultimately lies with employers.

In cases of sonographers without appropriate professional registration, employers will normally require educational qualifications such as a Masters or Postgraduate certificate (or diploma) in medical or clinical ultrasound from a University operating a course accredited by the Consortium for the Accreditation of Sonographic Education (CASE). These courses usually take 12 to 18 months to complete and

include academic study as well as clinical placement at an approved hospital. Three year undergraduate courses in ultrasound are also accredited by CASE at some Universities. After completing such courses graduates will expect to work as sonographers and will continue to gain experience under the supervision of their employer, a healthcare provider, who will also manage their training and career progression as well the quality of their work.

The Royal College of Radiologists (RCR) and Society and College of Radiographers (SCoR) have published standards for the provision of an ultrasound service⁹⁹ including requirements for training and education. These standards acknowledge the importance of the specialist skills and knowledge of the operator in the delivery of ultrasound services. They also specify minimum levels of qualifications, the support needed by newly qualified sonographers and requirements for continuing professional development such as attending training courses, conferences and so forth. However, providers of medical ultrasound services in England are not obliged to adopt these standards. Instead there is a legal requirement for them to be registered with the Care Quality Commission (CQC) and this registration depends on passing periodic audits to demonstrate compliance with core CQC standards as well as compliance with relevant regulations and statutes, such as the Health and Social Care Act 2008⁸⁵. Therefore sonographers are subject to QA process defined and operated by their employers, but one which is audited and approved on a statutory basis by the CQC.

Quality Assurance of Ultrasound Equipment

The provision, operation and maintenance of medical ultrasound equipment by healthcare providers in the UK are subject to legislation such as the EU Directive on Medical Devices 1993¹⁰⁰. Healthcare providers are audited by the CQC to assess performance against standards which take account of such statutory requirements. Consequently ultrasound equipment is subject to quality assurance with the aims of ensuring it is safe, properly maintained and suitable for the way in which it is being used; objectives also supported by the National Health Service Litigation Authority (NHSLA)¹⁰¹. However, as reported by Dudley et al¹⁰², healthcare providers each implement such QA programmes with varying degrees of formality.

Russell¹⁰³ draws attention to the degradation of ultrasound machine performance even in the case of modern devices with digital circuits capable of self-adjustment and recalibration. These are problems which may lead to image quality faults that are clinically significant, as reported by Dudley and Gibson¹⁰⁴. There is also, as reported by Dudley, a need for the machine operator to perform routine cleaning, inspection and maintenance. Therefore the QA of ultrasound machines requires regular work at both a technical and user level. The proper recording of the quality checks and maintenance performed on each machine is an important part of QA in both of these areas as it provides a structure for the work as well as evidence for any subsequent audit.

Technical Level Quality Checks

At a technical level quality checks will typically be carried-out by a Medical Physics Expert (MPE) employed by the healthcare provider, possibly working in conjunction with the third-party technicians associated with the equipment manufacturer of specific machines. This type of work covers the commissioning of new machines as well as their subsequent inspection and maintenance which is typically performed monthly and annually (or biannually) with corresponding levels of thoroughness. These checks start with simple visual assessment of electrical and mechanical integrity of the equipment, including probes and cable. At a more detailed level checks are made for scan line drop out as well as the grayscale reliability of monitors and printouts. The more complex tests involve measuring changes in sensitivity and noise using calibrated test objects as well as checking calliper accuracy and precision. The use of the resolution integral pioneered by Pye and Ellis³⁶ and described in the context of the Edinburgh Pipe phantom by Moran et al³⁵ may also be used to quantify faults in ultrasound probes or machines as well as to evaluate technical performance for sensitivity and resolution as described in Chapter 2. The results may then be compared against standards such as those set by the Royal College of Radiologists¹⁰⁵.

Dudley and Griffith¹⁰⁶ report on the limitations of checking noise and sensitivity using the testing methods for ultrasound machines recommended by two professional bodies. They present alternative simpler tests in an attempt to improve the detection of faults causing a deterioration of clinical performance. In later work

Dudley and Gibson¹⁰⁴ report (2014) on the problems of performing technical quality checks using manual measurement and visual assessment of images generated from test objects. They suggest, as an alternative, a computer system able to measure imaging performance parameters in an objective, verifiable and repeatable way. Therefore it would seem that the QA of ultrasound machines at an advanced technical level is still a significant challenge and there are different views on the best approach.

The key to the effective QA of ultrasound equipment at any technical level is the recording of the checks and maintenance performed over its lifetime. Typically this would involve keeping a log book for each individual machine, probe and cable. Clearly, a computerised system has advantages over paper in terms of data security and ease of audit when many sets of ultrasound equipment are being managed. However, published information about such equipment logs is not readily available.

User Level Quality Checks

BMUS provides guidelines¹⁰² for the regular QA checking and maintenance of ultrasound scanners at the user level. They attempt to unify guidance to sonographers from various sources about the QA work they should perform on their equipment. The guideline cover three levels of QA; 1) infection control and scanner damage 2) basic scanner and transducer testing 3) further scanner and transducer testing. However, these are guidelines rather than regulations, so their implementation will depend on the QA programme of individual healthcare providers.

The BMUS guide suggests that infection control (issues detailed by Miyague et al¹⁰⁷) should be performed before scanning each patient whereas checks for scanner damage should be done weekly. However, before using any equipment on a patient the attending medical professional has a duty to take reasonable steps to ensure it is safe and functional. Therefore, though not explicitly stated by the guide, it would seem prudent for the sonographer at the start of each shift to perform many of the level 1 checks, such as inspecting the probe, cables and machine for any obvious damage or wear. Other level 1 checks, such as inspecting the air filters, might be performed less regularly.

Level 2 checks cover the general operation of the equipment and the BMUS guide suggests that sonographers should perform them daily. Again, although not stated, it would seem prudent for a sonographer to check the condition of the room as well as the equipment before attempting to scan the first patient. Checking the room should include general cleanliness, temperature and lighting levels as well as the provision of supplies like infection control materials and furniture such as the scanning chair and operator stool. The guidelines are more specific in terms of checking the equipment. It is suggested that the brightness and contrast of the monitor should be adjusted so a greyscale bar is appropriately displayed. In addition the sonographer should operate the machine with the probe in free air to check the image for dropout or lack of uniformity. Dropout is a loss of echo at certain points in the X-axis of the reverberation pattern that appears in image when the probe is operated in free air; see Figure 3-3. It usually occurs when a few elements of the probe have failed, but it can also be caused by cable failure or electronics failure in the front-end electronics (see Chapter 2). Lack of uniformity in the reverberation pattern indicates similar problems as drop-out, but suggests degradation rather than complete failure of a channel.

The BMUS guide proposes that level 3 checks should be performed each month. These checks involved a more detailed inspection of the equipment for problems relating to sensitivity, noise and dropout. It also requires inspection of the machine's electrical safety label to check that the next test is not due within the next month. A check for sensitivity is typically performed by operating the machine with the probe in free air, selecting baseline settings for gain etc., and then measuring the distance from the top of the image to the deepest visible reverberation in the centre of the image; see Figure 3-3. This measurement should be compared with the value recorded for the previous month and any significant change reported. A check for noise is similar to the sensitivity check, but involves reducing the gain until noise disappears from the image. The gain setting at this point is compared to the value recorded for the previous month and again any significant change is reported. The dropout check at level 3 is slightly more detailed than the check at level 2 and involves passing a paper click along the probe face to produce strong localised echoes which allows a better assessment of potential problems.

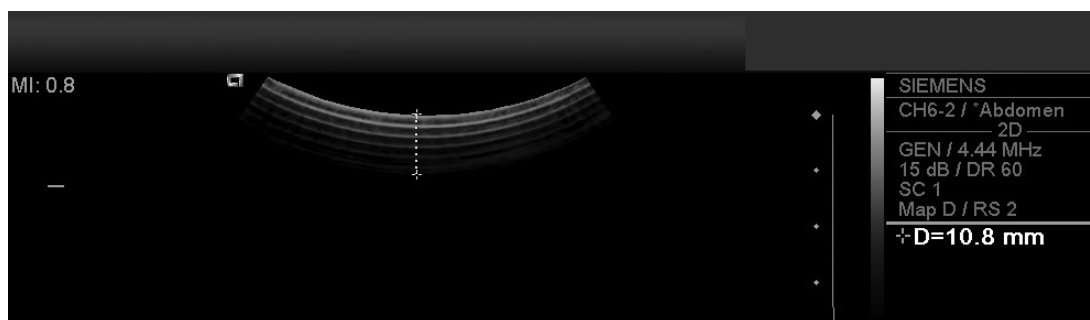


Figure 3.3: Reverberation pattern produced by operation of the probe in free air. Any reduction in the sensitivity of the machine can be identified by a change in the distance between the reverberation lines (measured in this instance as 10.8mm). Image ©Steve Pye.

The BMUS guidelines do not suggest how the various checks should be recorded. However, it would seem sensible to record the allocation of equipment to particular rooms as well as the staff who have used the room for scanning at a given time and date. In this way the discovery of an equipment fault with clinical significance can trigger the recall of patients who might be impacted, assuming a patient list is also kept with details of who examined them and when. In addition, it would be clearly beneficial to keep a log for each set of equipment (probes, cables and machine) to record the weekly and monthly checks performed by sonographers as well as requests for maintenance. This will help ensure that the QA process is followed, particularly in the case of equipment shared by multiple sonographers. It will also provide evidence for audit purposes. However, published information about the recording of user level QA for ultrasound equipment is not readily available, so it is presumed that there is no common standard within the NHS or other large UK healthcare providers.

Quality Assurance of Ultrasound Procedures

Professional bodies such as the Royal College of Radiologists (RCR) and the Royal College of Obstetricians & Gynaecologists (RCOG) provide guidance about ultrasound procedures at a general level as well as for specific types of examinations. This guidance, together with that provided by BMUS, may help specify the QA processes adopted by healthcare providers in respect the actual ultrasound examinations performed on patients. However, the adoption of the guidance is fragmented because the actual QA process is defined and operated by individual healthcare providers.

General Considerations for the QA of Ultrasound Procedures

It is relatively easy to define QA processes for certain aspects of an ultrasound examination, particularly those that apply to most types of procedures. For example, as previously mentioned, at the start of the examination standards can be set for the scanning environment like room temperature, lighting and furniture. It is also not difficult to define procedures for identifying a patient, specifying the information that must be provided about them in terms of the justification and objectives of the examination. Similarly, a standard approach can be taken to explaining risks, obtaining informed consent and addressing cultural issues such as language barriers, chaperone requirements and so forth.

During the ultrasound procedure itself a number of QA processes can also be defined to ensure the safety of both patient and sonographer. Patient safety QA includes provision for adequate infection control as well as making sure the power and duration of exposure to the ultrasound beam within acceptable limits for TI and MI (see Chapter 2). Sonographer safety QA covers issues such as repetitive strain injury (RSI) or back pain resulting from badly adjusted seating.

After the examination is complete, QA processes will help ensure that the report is written in an appropriate way and contains standard pieces of information needed for the effective future treatment of the patient. The adoption of report forms can help not only in terms of providing a structure for the execution of the process, but also to facilitate its subsequent audit. In addition, the static images recorded from the ultrasound machine's 'cine loop' (see Chapter 2) will also need to be recorded and stored in a PACS system in a standard way to facilitate future retrieval and inspection; again a process that is easy to specify and control.

It is not surprising that the above general considerations for the performance of ultrasound procedures should be well controlled by QA processes. They are more suited to the setting and control of standards than the specifics of particular procedures, such as detecting ovarian abnormalities – a task highly dependent upon the operator's skill and experience² so correspondingly much more difficult to measure and control.

Quality of Specific Ultrasound Procedures

In the publication ‘standards for provision of an ultrasound service’ (RCR¹⁰⁵) emphasis is given in section 6 to the difficulties of providing robust QA for non-obstetric ultrasound imaging. It is pointed out that ultrasound is unique in terms of being the only imaging modality that combines image assessment and diagnosis. That is to say the value lies in the examination, not in the static images it might record. It is accepted that such images are only representative of the ultrasound examination and not a complete record. Furthermore *‘one of the great strengths of ultrasound is the ability to image anatomical structures in real-time, in a variety of different planes, using a variety of machine settings to optimise visualisation of anatomical and pathological structures’*. It is suggested therefore that attempts to enforce a rigid procedure for ultrasound examination performed as a diagnostic are likely to be counter-productive. However, it is also accepted that it is possible to perform standardised, protocol-driven ultrasound imaging in some cases. The time allowed for each type of procedure may also be standardised, though the BMUS guidelines make clear the need for allowing enough time for a competent examination irrespective of management goals.

Ultrasound procedures performed for the purpose of diagnosis vs. screening

Cantin and Knapp¹⁰⁸ likewise consider the problems of providing QA for ultrasound procedures and discuss the differences between performing an examination for diagnostic purposes and performing it as a screen for a condition such as prenatal detection of Down’s syndrome. It is suggested that ultrasound procedures for screening usually involve the production of a few standard images to measure specific features whereas procedures for diagnostic purposes may need to produce large numbers of images, often in non-standard planes or angles. This reflects the difference between a procedure given to identify someone at a higher risk of a health problem (screen) and a procedure intended to provide firm evidence of a health problem (diagnostic). Therefore setting audit criteria for procedures related to ultrasound screening is more straight-forward than for procedures that are diagnostic in nature. It is for this reason that QA for specific ultrasound procedures given for diagnostic purposes need to be considered separately from those given as a screen for a particular condition.

Diagnostic ultrasound procedures

BMUS issues professional guidance for professional ultrasound practice¹⁰⁹. This is a 'live' web-based document based on previous work in 2008 by SCoR members which has the aim of providing up-to-date practice guidance to sonographers. Most of the document is concerned with giving advice about particular diagnostic ultrasound procedures. The procedures cover the main applications of diagnostic scanning; gynaecological, abdominal (liver, gallbladder, pancreas, uro-genital), head and neck, musculoskeletal (shoulder, elbow, wrist and hand, hip, knee, foot and ankle), paediatric and neonatal.

Typically, guidance for a particular procedure covers the clinical history that should be provided, techniques that are appropriate (for example TVS requires an empty bladder), the structures that should be examined (for example ovary echogenicity, position, size, shape, follicles) and details of specific pathology (for example, a simple cyst appears well defined, echo-free, uni-locular, etc.). In addition other sources may be referenced, for example NICE guidelines for the recognition and management of ovarian cancer¹¹⁰, or International Ovarian Tissue Analysis (IOTA)¹¹¹ rules for characterising ovarian masses.

The RCOG 'Green-top' guides¹¹² have a similar purpose to those published by BMUS, but have a more formal structure and give less practical advice about ultrasound technique as they are more focused on applicable diagnostic procedures and disease management. The Green-top guides, like the BMUS guidelines, have relevance to QA only in terms of allowing a specific diagnostic procedure to be specified as following their advice. These guides do not in themselves seek to set standards for which an ultrasound procedure can obtain compliance.

Approaches to the quality control of specific diagnostic procedures typically involve a retrospective review of the static images produced by the sonographer as evidence to support her report. However, RCR standards¹⁰⁵ also suggest that an audit may involve an expert repeating a scan during the same patient visit to provide comparative results, or a follow-up of the outcome for a collection of similar cases.

A retrospective review of the static images from an ultrasound exam may be performed by one or more experts in the field (see Chapter 8, 9) or it may be

performed by a group of the sonographer's peers as reported by Cantin and Knapp¹⁰⁸. These sorts of reviews usually involve the completion of a questionnaire to help ensure the uniformity of scoring criteria. A number of standard questionnaires exist, commonly referred to as 'audit tools' and produced in Microsoft Word or Adobe PDF format. The questionnaire used by Cantin and Knapp contained nine questions given in a variety of forms; yes/no answers (binary), discrete scoring levels such as poor, acceptable and good (3-point Likert), and making a mark on a line (continuous scale). Similar questionnaires are proposed in the RCR standards for scoring both ultrasound images and reports.

QA processes based on auditing the sonographer's work by review can require significant resources. For example, 9 experts in gynaecological scanning collectively spent 150 hours performing an audit to confirm self-reported ovarian visualisation rates on just seven sonographers; see chapter 8. Therefore, such audits are often precluded on the grounds of cost or the availability of suitable experts. Instead it is common practice to ask sonographers to collect their own QC metrics from their work. However, such self-reporting can deliver unreliable QC data and so subvert the outcome of the QA process; a problem this thesis seeks to address, as described in chapters 7, 8 and 9.

Ultrasound screening procedures

The UK National Screening Committee (NSC) identifies four health screening programmes that include ultrasound examination:

- Fetal Anomaly Screening Programme (FASP)
- NHS Breast Screening Programme (BSP)
- NHS Abdominal Aortic Aneurysm (AAA) screening program
- NHS Neonatal and Infant Physical Examination (NIPE) screening programme.

FASP¹¹³ has particular relevance to any future screening programme for ovarian cancers using TVS scanning as it addresses many of the QA issues associated with performing an ultrasound examination as a screen for a health condition as opposed to a diagnostic, so it provides the basis for discussion. FASP performs a combined screen test for Down's (T21) syndrome. It requires the woman to attend a clinic at

between 10 and 14 weeks of pregnancy where an ultrasound examination is conducted to measure her foetus' crown-rump length (CRL) and nuchal translucency (NT). The values are entered into a report and sent to a laboratory together with a blood sample. The laboratory determines values for various blood serum biomarkers and then inputs all the data into a certified computer program¹¹⁴ to calculate risk of T21 in the foetus, adjusted for factors like maternal age and the presence of twins. Using these methods FASP report¹¹⁵ a T21 detection rate of approximately 80% with a false positive rate of 3%.

CRL is the distance between the top of the head and bottom of the buttock and allows calculation of the gestational age (within certain limits). NT is the thickness (mm) of a subcutaneous collection of fluid in the fetal nuchal region (nape of the neck) and is a useful age related biomarker for Down syndrome, as demonstrated by Nicolaides et al¹¹⁶. For the screen to give a reliable result both CRL and NT need to be measured precisely.

FASP publishes¹¹⁷ detailed specifications for measuring NT and CRL as well as prescribing the QA processes¹¹⁸ to ensure national standards¹¹⁹ are maintained for the delivery of screening to pregnant women. These processes are monitored and audited by the NHS sponsored Down's syndrome screening Quality Assurance Support Service (DQASS) with the aim of continuously improving the quality of screening through feedback to sonographers and their departments as well as to laboratories and suppliers. DQASS work with Screening Support Sonographers (SSS) nominated by local screening providers who undertake the training and support of sonographers as well as tasks such as conducting departmental image reviews, record keeping, creating and implementing improvement.

Maintaining the precision of NT measurements depends upon processing the data effectively to minimise the impact of variation due to gestational age. Therefore the value is compared to a reference curve giving median NT values for CRL values so allowing adjustment for gestational age. The reference curve used by FASP for T21 screening is produced by the Fetal Medical Foundation (FMF)¹²⁰ from studies of NT and CRL values for a population, though other reference curves may be used for QA purposes. This allows the absolute value of NT thickness to be analyzed using the

method of multiples of the median (MoM); measured value divided by the median of the population.

FASP achieves quality control for NT and CRL measurements by monitoring datasets of dimensions produced over a time period by different local screening providers. This allows a collection of measurement to be compared with the reference curve in order to detect bias; see Figure 3-4. A system of flags is used to provide feedback and prioritise training such that a green flag is reported when the bias detected is less than or equal to 0.10mm and a red flag is reported when the bias is greater than 0.4mm. It is recognised that bias may arise due to a) machine and environmental factors, b) human factors associated with the sonographer and scanning department, c) demographic factors such as high prevalence of women with elevated BMI in a particular district.

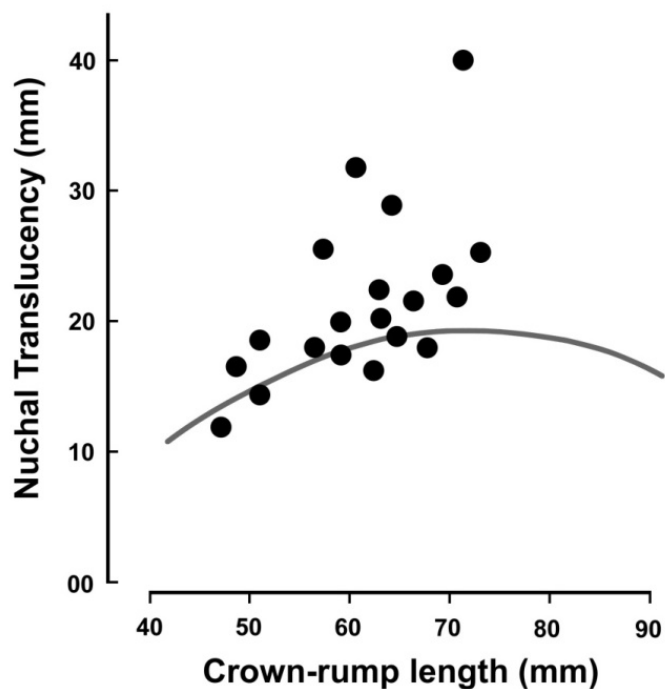


Figure 3-4: NT and CRL Quality Control Reference Curve. The curve in the centre of the chart shows the expected range of crown-rump length to nuchal translucency size for a group of sonographers. The dots are mean results for a particular sonographer from a large collection of measurements. The chart suggests that the sonographer with the outlying result (NT=40, CRL=72) may not be measuring the dimensions correctly.

The American College of Medical Genetics (ACMG) has produced detailed standards and guidelines for T21 screening in the form of a paper by Palomaki et al¹²¹. Emphasis is given to the importance of maintaining screen quality through

sonographer training, image review and QA of the quantitative NT and CRL measurements. It is suggested that sonographers should have paired NT/CRL measurements from at least 30 pregnancies validated against standard reference curves before performing clinical screens with on-going monitoring of their data to confirm variation is maintained with expected limits. Hynek et al¹²² present an alternative approach to NT quality control based on the statistic process control (SPC) methods discussed earlier in this chapter.

Ultrasound screening procedures for Ovarian Cancer

The NSC does not currently recommend systematic population screening in the UK for ovarian cancer¹²³. However, this statement may be altered in the future as a result of further findings by UKCTOCS, particularly in respect of the impact of screening on mortality during the extended follow-up period to 2018. At present research into tests that might be used in a national screening programme is focused on the CA125 blood test and TVS examination.

A search for articles was made in the Web of Science database on 1st March 2016 for articles specific about quality control of Transvaginal Ultrasound (TVS); search terms topics “transvaginal ultrasound quality”, dates 1990 – 2016. There was no intention to publish the results as a formal meta-analysis so the requirements of the PRISMA^b statement were not considered. A number of studies had assessed their methodological quality in respect of QUADAS^c (quality assessments of diagnostic accuracy studies), but did not describe the QA approach for the diagnostic being studied so were excluded. The search produced 355 results of which 5 described the details of some form of quality procedure for TVS after a review of the abstracts.

^b PRISMA (2009) revised the previous QUOROM statement (1996) aiming to improve the suboptimal reporting of meta-analyses. It defines the methodology of the review in terms of identification, filtering, eligibility and inclusion. It also has a checklist specifying what the review needs to report in respect of issues like its information sources, search strategy and so forth.

^c QUADAS-2 (2011) revised the previous QUADAS (2003) tool to assess the quality of diagnostic accuracy studies by assessing the risk of bias and applicability in 4 domains; patient selection, index test, reference standard, work flow and timing.

- Sharma et al² reported on the QA approach adopted by UKCTOCS ovarian cancer screening trial.
- Weissfeld et al¹²⁴ reported on the QA approach adopted by the Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening trial.
- Salomon et al¹²⁵ proposed a score-based method to improve the quality of ultrasound exams performed during obstetric emergencies. The stated aims of the study were 1) propose criteria for standardised image assessment 2) to evaluate the feasibility of implementing an image-based scoring method to produce QC metrics from such criteria 3) to evaluate the impact of training on the QC metrics.
- Bredella et al¹²⁶ reported on a QA programme involving experts at a central location independently measuring endometrial thickness from images collected from 2,000 TVS exams given as part of a multi-centre drugs trial. The stated quality criteria was image quality sufficient for endometrial measurement by an expert and measurement by expert agreeing with sonographer recorded measurement with $\pm 2\text{mm}$. The examinations were repeated if found unacceptable and feedback was given to individual sonographers by the experts in order to improve future performance.
- Burger et al¹²⁷ reported on the impact of quality control standards on reducing inter-observer variability in cervical length measurements using TVS images. At the start of the study measurements of cervical lengths were made, QC standards were then established and their impact on inter-observer variability was measured.

It can be concluded that the most significant work done of the QA of TVS used as a screen for ovarian cancer has been performed by the PLCO and UKCTOCS, with useful contributions made by Salomen et al in terms of an image-based scoring method to produce QC metrics. The work by Bredella et al and Burger et al in proposing ways of measuring inter-observer variability is applicable to measuring endometrial thickness and cervical length respectively, so has limited value in terms of screening for ovarian cancer.

Quality Assurance of TVS in Ovarian Cancer Screening Trials

Four major trials have been conducted to evaluate the efficacy of TVS as a screen for ovarian cancer. The dates for commencement of enrolment, locations and names are:

- 1985, Japan: Shizuoka Cohort Study on Ovarian Cancer Screening (SCSOCS)
- 1987, USA: Kentucky Ovarian Cancer Ultrasound Screening Study
- 1994, USA: Prostrate, Lung, Colorectal and Ovarian (PLCO) screening trial
- 2001, UK: Collaborative Trial of Ovarian Cancer Screening (UKCTOCS)

Papers about the QA processes for TVS scanning adopted by these trials have been published for UKCTOCS² and PLCO. No information about the QA of the earlier trials has been published to date. This may reflect recognition in the later trials of the importance of QA in large scale screening.

Quality Assurance of TVS in PLCO

A number of papers have been published about QA in the PLCO trial. Gohagan et al¹²⁸ described the trial's overall approach to QA and its organisational structure as well as detailing its site visit/audit process and IT systems. Weissfield et al¹²⁴ described the quality control plan and procedures for the various examinations performed during the trial, including TVS for ovarian cancer screening. Hasson et al¹²⁹ described the design and evolution of the Data Management Systems and gave details of the quality assurance approach, whilst specific details of the computer systems employed to provide QA for imaging were given by Moore et al¹³⁰ (though only in the context of Computed Tomography (CT) and chest X-ray for lung cancer screening).

During the early planning of the PLCO trial it was recognised that “*the credibility of findings would depend on the quality of [its] design and execution*” (Gohagan et al). Therefore well before screened started, QA was given a high priority and considerable effort was made in designing and implementing processes to ensure the quality of the work. This effort continued for more than twenty years and was described by Gohagan et al as being “*operationally analogous to the concept of Total Quality Management (TQM)*”. There is extensive documentation covering all

aspects of the trial and its approach to quality from governance and selection of collaborators¹³¹, through the delivery of screening, to the analysis and presentation of the data¹³².

Local QA plans were produced by each screening centre (SC) during the PLCO pilot phase which were reviewed and approved by the Coordinating Centre (CC) to ensure consistency and compliance with the trial objectives. Specifications were also created for the examiners, equipment and procedures which were codified into a common Manual Of Operations and Procedures (MOOP). In addition, a QA subcommittee was formed to coordinate and oversee quality in terms of the setting and monitoring of standards as well as making changes to improve performance (Plan-Do-Check-Act cycle). This subcommittee also led the work on creating a culture of quality and continuous improvement which over several decades evolved to become a methodology similar to TQM.

The PLCO Quality Subcommittee published a formal QA plan with the aims of standardising examination procedures and maintaining accountability. It set standards in terms of definitions, data collection, policies and procedures. It also defined how these standards should be monitored in terms of periodic analysis of screening test data, independent repetition or review of examinations and site visits to observe test performance. Furthermore it specified how examiners and their trainers should be centrally certified and registered, making a clear distinction between the two roles with the aim of ensuring that they were performed by separate people with appropriate qualifications. In the case of TVS, for example, PLCO certified examiners were required to be ARDMS-registered technicians who had both passed the OB/Gyn section of the certification examination and also performed 50-100 TVS exams.

Details about the QA of ultrasound equipment were not found any of the papers published by PLCO and it is unclear whether the recommendations of Dudley et al were observed; see previous section. However, papers were published about the IT systems developed for the trial and their role in QA:

- Hasson et al¹²⁹ state that the system design was guided by the Federal Information Processing Standard (FIPS) and the Yourdon Structured

Method¹³³ was used to direct development activity. This suggests a software team was employed to design, implement and test the systems in a regulated fashion. Hasson et al also describe in some detail the QA adopted to ensure the integrity of the data. This included employing bar-codes and techniques such as double entry or check-sum validation to reduce data entry errors. It also including assigning ownership to each data object so each SC had responsibility for all the data it collected and a subsequent change could only be requested by the Coordinating Centre, as the change itself had to be implemented by the SC. Therefore all data entry and any subsequent editing was controlled and audited with regular reports being made by the SC to the Coordinating Centre in this regard so that poorly performing centres could be identified and improvement action taken.

- Moore et al¹³⁰ describe an IT system to allow QA of the images for the National Lung Screening Trial (NLST) which was conducted within the administrative organisation of the PLCO as well as the American College of Radiology Imaging Network (ACRIN). This allows SCs to submit image studies over the internet to a Quality Assurance Coordinating Centre (QACC) according to a sample list compiled using data from the PLCO trial. This system had functionality similar to the UKCTOCS URA (chapter 4) in terms of anonymising and storing DICOM images, but also allowed these images to be transferred to a web-based system (EasyWeb¹³⁴) for review by experts at multiple locations. Features were provided for processing such image studies as a workflow and generating reports detailing the results of the review. However, the system was only used for purposes of QA for CT and chest X-Ray images used in screening for lung cancer.

The QA of the examination procedures was based on standards set in the MOOP. Weissfield et al describe how the quality of exams (including TVS) was monitored in respect of these standards:

- Proportions of test-positive (prompting referral for diagnostic evaluation) and test-inadequate (screens performed defectively) were monitored and analyzed by examiner, by scanning centre and by examination sequence with a baseline created for the first exam soon after randomisation and subsequent

exams being performed on annual basis. This allowed the identification of examiners and centres that appeared to be performing in a non-standard way.

- Proportions of TVS exams were monitored and analyzed by examiner and by scanning centre in respect of having visualised one or both ovaries or having reported ovary abnormalities. Mean ovarian volume reported by TVS exams was also monitored and analyzed. This allowed the identification of examiners and centres that were producing results outside the expected ranges for such exams.
- Inter-observer variability was monitored in respect of a sample of exams that were repeated at the same participant visit. In the case of TVS, 164 exams were repeated each year for each SC so inter-observer variability could be compared and outliers identified.

The results of the monitoring were used to create reports to help target improvement at particular SCs as well as for individual examiners. Other types of examination reviews were also performed, including reviews of TVS static images similar to those described in Chapters 7 and 9. However, such reviews did not form part of the formal QA process as they were only performed on an irregular basis and the results were not used to target improvement.

Quality Assurance of TVS in UKCTOCS

UKCTOCS created a document in 2006 about its approach to Ultrasound Quality Assurance which is reproduced in Appendix E. In addition the standards for TVS scanning are defined in the trial protocol, the relevant sections of which are given in Appendix B. The following discussion is based on this information, published papers as well as some documents and presentations made available for the purposes of the thesis.

Published Papers

To date the only paper published specifically about QA for TVS scanning in UKCTOCS was written by Sharma et al², though a paper was published about the acceptability of TVS amongst volunteers (Sharma et al¹³⁵) and a further paper was published about factors influencing visualization rate (VR) of ovaries (Sharma et al¹³⁰).

Much of the QA work described by Sharma et al² relates to the adjustment of the ovary visualisation rate (VR) metric for bias due to differences between the sets of volunteers scanned by different sonographers in the thirteen trial centres like age, Body Mass Index (BMI) and so forth. It concludes that subjective factors such as individual skill in scanning, attention to detail and experience are the major contributors to variation in VR, so unadjusted VR values could be used for making comparisons.

UKCTOCS Quality Assurance of TVS Scanning

The approach taken by the trial to the quality management of TVS Scanning is documented in Appendix E. It summarises the work done by UKCTOCS prior to 2006 in respect of quality assurance for the scanning protocol, audit requirements, equipment, data collection, staff selection, training and communication. The document also sets out the plan for QA work from 2006 onwards including:

- Providing individual feedback to all sonographers every six months including details of performance using data from their scan report forms and details of any screen negative examinations they had completed in respect of women who had been subsequently diagnosed with ovarian cancer together with comments arising from a review of the associated images by the Co-investigator.
- Instigation of annual visits to trial centres by members of the Ultrasound Subcommittee (USC) formed earlier in the year.
- Setting-up of an accreditation and certification programme for sonographers.

The USC was responsible for the quality management of TVS scanning. However, unlike PLCO, it did not attempt to implement best practice from a formal quality methodology, such as TQM. This probably reflects the differences of approach to quality in healthcare between USA and UK at the time.

The Ultrasound Subcommittee (USC) reported to the Trial Management Committee, the group ultimately responsible for the trial. The USC was chaired by the Principal Investigator / Trial Coordinator and its members included the Co-Investigator with core expertise in ultrasound scanning, members of the Coordinating Centre team

(CC-US) involved in the implementation of ultrasound screening as well as various senior radiologists and sonographers from the trial centres who were directly involved in the delivery of TVS scanning. A key member of the CC-US team was the Clinical Fellow whose remit was to oversee the ultrasound implementation on a day to day basis, work on QA and fail safe monitoring in the USS arm. Others on the CC-US team included the Trial Manager, the Data Manager, the Trial Statistician and the Senior Nurse, all of whom liaised with the Trial Centre teams on a daily basis.

The USC was responsible for oversight of the ultrasound arm of the trial as well as the CC-US group. In this regard the USC set strategic objectives of the CC-US, monitored its work and took action as needed to ensure the ultrasound arm of the trial was operating correctly. For example, it initiated the audit described in Chapter 7. Members of the USC with expertise in gynaecological scanning also performed technical work such as reviewing images from scans (particularly those of women with interval cancers^d), inspecting Trial Centres or assessing individual sonographers, particularly for those achieving low rates of ovary visualisation. The Co-Investigator member of the USC was particularly active in this regard.

All Trial Centre sites were visited by the Clinical Fellow and the Co-Investigator in 2006 and a detailed report about each one was produced which included details about the quality of sonographer scanning which had been observed. The Clinical Fellow also worked with the Trial Statistician to produce reports for the USC meetings which were held approximately every six months after 2006, though information was also communicated in the intervening periods, typically by email.

The USC decided in 2006 that the monitoring and support of the sonographers might be better achieved if it was provided by a senior member from within their own profession. Therefore in 2007 the USC appointed a senior ultrasonographer as the UKCTOCS National Lead Sonographer (NLS) to take over the duties of the Clinical

^d Initial screening of volunteers revealed a number of cancers considered as disease that had existed before the start of the trial, so were termed prevalence cancers. Interval cancers were detected after this initial screen.

Fellow and lead on day-to-day management of scanning for the trial. The NLS was a superintendent level sonographer who had extensive experience in managing a team of sonographers as well as significant expertise in gynaecological scanning, especially TVS.

Development of Standards and Accreditation

A key contribution made by the NLS was the introduction of a system of sonographer accreditation which was analogous to the PLCO sonographer certification reported by Weissfeld et al. Therefore from 2008 all UKCTOCS sonographers were subject to annual accreditation to confirm they were operating according to the standards set in the Standard Operating Procedure (SOP). This document served a similar purpose to the MOOP created by PLCO in terms of setting quality standards which could be monitored as part of a QA process. Sonographer accreditation involved the completion of a questionnaire to show an understanding of the trial protocol and SOP, review of Visualization Rates (VR) achieved during the previous three months with a requirement to achieve visualisation of one or both ovaries for at least 60% of their exams, practical assessment by the NLS of scanning performance in respect of at least five TVS examinations at the sonographer's Trial Centre, submission of images from nine examinations to Coordinating Centre for review by NLS and a USC committee member having gynaecological scanning experience.

Sonographers who failed to achieve the standards set by the accreditation process were not allowed to perform further TVS scanning for UKCTOCS, though training and support was provided to help such sonographers reach the required standard. Accreditation was also used to direct subsequent training for individual sonographers with the aim of improving their future performance. In contrast the PLCO certification process seems only to have been used during the selection of sonographers and had no role in ensuing quality improvement. Therefore UKCTOCS accreditation was an important innovation for large scale TVS scanning and formed a key part of QA for the ultrasound arm after 2008.

Fail-Safe Monitoring of Scanning Report Forms

The NLS continued the work of the Clinical Fellow by conducting ‘fail-safe’ monitoring of all examinations. This involved a weekly review of information in the free-text fields from the latest scanning report forms (see Appendix C). This was necessary because it had been observed that occasionally sonographers reported ovaries as normal or unsatisfactory, but also reported abnormal findings in these fields. Therefore there was a risk that the classification performed by the TMS algorithm might result in abnormalities in scans being missed. Fail-safe monitoring mitigated against this risk by assuming all exams were problematic until their reports had been reviewed by the NLS (or Clinical Fellow) – i.e. the reviewed process operated on a fail-safe principle. Approximately 500 exams per week were reviewed in this way.

Monitoring of Abnormal Scans and Image Reviews

In addition to fail-safe monitoring the Senior Nurse reviewed all information from the reports of scans classified by the TMS as abnormal to confirm the findings and check for errors. Furthermore any volunteer diagnosed with an interval cancer who had been previously screened negative by TVS examination was subjected to a detailed review. This involved a review of the static images associated with previous examinations in order to determine whether they contained any suggestion of a tumour. These reviews were performed by the Co-Investigator assisted by the NLS and their outcomes were formally reported. Such reviews could also be requested by the trial centre in the case of a volunteer thought to need further consideration and were usually performed within 2 weeks of the request.

Sonographer QA

Sharma et al² report on quality procedures adopted for the selection and induction of UKCTOCS sonographers as well as their subsequent accreditation by the NLS (as previously described). Onsite and central training of sonographers are also described, though this must be considered separate to the trial’s QA work as no standards or monitoring requirements were detailed except for the requirement for all

sonographers to attend at least one of the two central training programmes (study days) which were held each year.

Two levels of sonographers were selected for the trial as described in Appendix B. Level 1 sonographers were required to be certified, or trained midwives or doctors with experience in gynaecological especially transvaginal scanning. However, the origin of such certifications is unclear as sonography is not recognised as a profession by the HCPC (see previous discussion) and the nature of training and experience was not precisely specified. Although this suboptimal from a QA perspective, it was not considered important in terms of the operation of UKCTOCS because the sonographers were also employed by the National Health Service (NHS) so their competence was not in doubt. Level II sonographers were required to be experienced gynaecologists or radiologists or senior ultrasonographers (usually at superintendent grade in the NHS) with particular expertise in transvaginal ultrasonography. Again, the nature of the nature of training and experience was not precisely specified, but their competence was assured by NHS selection processes.

UKCTOCS sonographer recruitment process was changed after 2008 so that candidates had to submit a curriculum vitae giving details of their qualifications and scanning experience. Further assurance about the quality of sonographers was provided by making the trial centre leader responsible for the recruitment process which included ensuring that all new recruits had a detailed knowledge of the UKCTOCS SOP as well as the ultrasound protocol. Sonographers new to UKCTOCS were also supervised by the NLS (or someone of similar professional standing) for a minimum of two examinations before being permitted to scan independently on the trial. They were then required to obtain UKCTOCS accreditation within three months of starting work.

Sharma et al did not report that the inception of a formal accreditation process had resulted in a significant number of sonographers leaving the employment of UKCTOCS, so it can be presumed that though the selection process before 2008 was less standardised, it was still effective in terms of ensuring the competency of UKCTOCS sonographers. However, it seems likely that the introduction of annual accreditation did result in quality improvement in TVS scanning as discussed below; Scanning Process QA.

Equipment QA and Role of IT Systems

Details about the QA of the ultrasound equipment were not given by Sharma et al, but similarly to PLCO, this probably reflects the fact that concern about such matters has only recently been raised by organisations like BMUS due to the latest papers by Dudley et al¹⁰². Therefore it is assumed that no formal QA was implemented in respect of the UKCTOCS ultrasound equipment, though it is understood that the SOP contained recommendations about the setup and operation of machines. A service contract was awarded to the company responsible for supplying the ultrasound equipment to the Trial Centres, but this did not include QC monitoring as part of a QA or preventative maintenance programme. Accordingly the machines were repaired only when defects were reported by Trial Centre staff.

UKCTOCS commissioned the development of two IT systems; the Trial Management System (TMS) and the Ultrasound Record Archive (URA). In respect of functionality required by the ultrasound arm of the trial, the primary purpose of the TMS was to allow entry of data by Trial Centre staff from the examination form (see Appendix C) and its subsequent reporting. Whereas the primary purpose of the URA was to automate the copying of examination images received from the Trial Centres into a central database at the Coordinating Centre, as described in Chapter 4.

In addition to its primary functions of collecting trial data and reporting results, the TMS was also required to classify TVS examinations as described in Chapter 2 and schedule scanning appointments as described in Chapter 1. It also implemented the functionality required by the multi-modal arm of the trial. Initial work on the TMS was performed by a small team employed by a commercial software company. Although a formal software development methodology was not employed, the TMS proved reliable in operation throughout the trial and no major defects were reported. Unlike the comparable PLCO system, the TMS did not implement validation techniques like double-entry or checksums in respect of data entered from the ultrasound arm of the trial. An annual budget of £100,000 was allocated to software and system maintenance so frequent updates were made during the course of the trial.

The URA was developed entirely by a sole developer, but an Agile approach was adopted based on methods described by Stott and Newkirk¹³⁶. This included formal unit testing, functional testing and system testing. In addition, configuration management was implemented to control deployment into the production environment. The last version of the software entered production in August 2007 and the URA was operated without further change until the end of the trial with no defects reported. URA data entry was restricted to the correction of errors in the DICOM file headers resulting from volunteer details being incorrectly entered into the ultrasound machines at the Trial Centres. Validation was performed by the URA in respect of requiring this data to be consistent with data in the TMS for the given examination.

A backlog of exam images awaiting entry into the URA had built-up by 2007 such that in some cases there was a delay of a year or more between the image being created and it being processed by the system. Therefore the URA was not used for QA purposes. Indeed, during the course of the trial the URA was only used on one occasion for validating the quality of images. This involved auditing the exams performed by seven sonographers over one year, as described in Chapter 8. By way of comparison, the PLCO equivalent system to the URA had been designed to allow the QC of images by independent experts located remotely from the coordinating centre, though this feature was only used for reviewing images from the lung cancer part of the trial, as previously mentioned. However, it can be speculated that the availability of such functionality in the URA may have encouraged additional QA for UKCTOCS scanning. Certainly the provision of better technology support in the ultrasound arm of the trial would have allowed a more sophisticated approach to its quality assurance as suggested by the tools subsequently developed for the thesis project, as described in Chapters 5 and 6.

Scanning Process QA

Quality Assurance monitoring of the scanning process primarily involved analysis of the visualisation rate (VR) of individual sonographers over a three month period, defined by Sharma et al as number of exams with right ovary seen divided by total exams performed. Sonographers achieving less than 60% VR in the three months before their annual accreditation were not allowed to continue scanning for

UKCTOCS, as previously described. The NLS also regularly monitored VR for sonographers in order to target assessment and on-site training. In addition periodic reports of VR both by sonographer and by Trial Centre were circulated with encoding so only the lead staff in each centre could identify their own results and those of their sonographers. Various statistics were also included in the reports to help identify deviations from expected performance, for example ovary volume comparisons between Trial Centres.

The need to have provided additional quality control of TVS scan reports is suggested by the discovery after the end of the trial that dimensions recorded in the TMS as D1, D2, D3 did not correspond to the accepted ordering for an ovary, possibly because the examination report did not specify the field names. Similarly, after the end of the trial it was discovered that a significant percentage of sonographers were not following SOP standards in terms measuring the longitudinal and transverse sections of the ovary or using the correct format for these images. These problems were identified as a result of the study described in Chapter 7 which in turn was dependent upon the tools developed for this thesis project.

An important finding by Sharma et al² is that the VR does not need to be adjusted to take account of variance between groups of volunteers being scanned at different Trial Centres for factors found to influence ovary visualization, as identified in their earlier study³⁰; age, age at menopause, hysterectomy, oophorectomy, tubal ligation, body mass index. The authors concluded that factors such as individual skill, attention to detail and experience are major contributors to differences in sonographer VR. Sharma et al² also state that “*ovarian visualisation is key to identifying morphological abnormalities that may be indicative of ovarian cancer*”. This suggests that an effective screen for ovarian cancer requires the visualisation of both ovaries (when present).

The VR results presented by Sharma et al² in respect of individual Trial Centres as well as their median are given in a graph which is reproduced as Figure 3-5 with the addition of lines to show two key events in the timeline; appointment of the NLS in 2006 and the introduction of new ultrasound machines in 2008. The largest improvement in median VR happened a year after the appointment of the NLS and during this time all Trial Centres apart from one reported improvements. It should be

noted that this one exception had reported the highest VR at the start of 2006 and the reduction was only modest. It seems unlikely that the training initiated by the NLS would have such an immediate impact in its first year, particularly as accreditation did not commence until 2007. It seems more likely that VR improved as a consequence of sonographers being made aware of its importance, particularly by such a well-respected and senior colleague. This improvement may have been caused by sonographers performing examinations more diligently, or by an increase in the over-reporting of VR (Chapter 8 and 9), or by a combination of these and other factors.

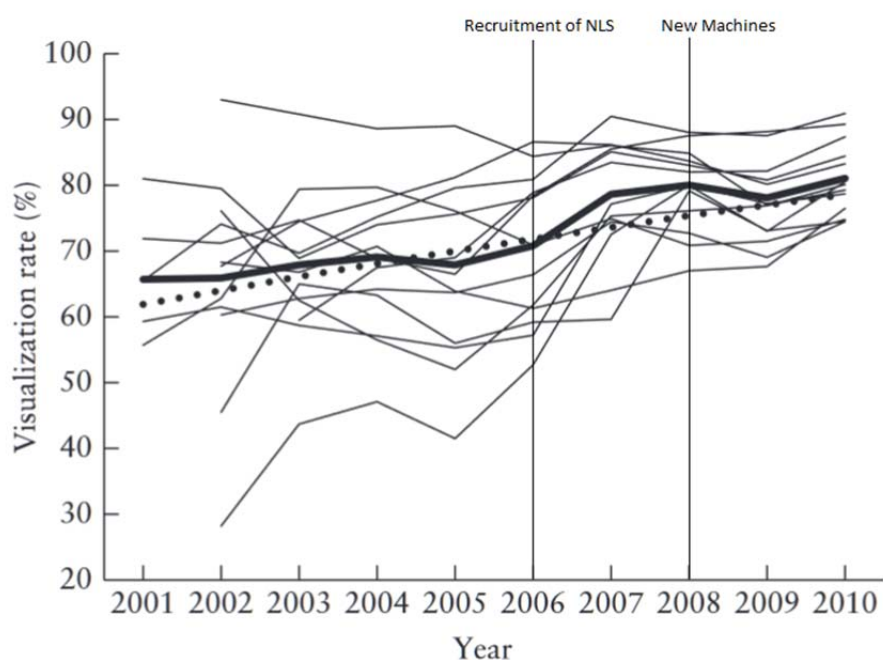


Figure 3-5: Annual Visualisation Rates (VR) for right ovary arising from TVS examinations performed on 48,250 postmenopausal women in the 13 UKTOCS Trial Centres. The thin lines show median VR for individual Trial Centres and the thick line is median VR for all centres. The dotted line shows media VR for all centres with adjustment for non-subjective factors that influence visualisation. © 2015 A. Sharma et al².

The reduction in VR observed in almost all Trial Centres after the introduction of new ultrasound machines at the start of 2008 may reflect the supposition made in Chapter 2 that better machines do not necessarily result in improvements in ovary visualisation, though undoubtedly better resolution does assist in the diagnosis of abnormalities once the ovary is located. Indeed, sonographers may have found the new machines initially more difficult to operate which would explain the widespread reduction in VR in 2009. An alternative explanation may be that the audit of seven

sonographers performed in 2009 (Chapter 8) caused a reduction in over-reporting of VR amongst other sonographers due to the realisation that UKCTOCS had the capability to conduct such audits in the future.

VR is described by Sharma et al² as an imperfect metric as it does not include scans which have visualised the left ovary, but not the right because it had been removed by previous oophorectomy. It is also described as imperfect as it does not include scans which have not seen the right ovary, but have obtained a ‘good view’ of the pelvic sidewall (the trial protocol actually defines such a ‘good view’ as visualisation of the Iliac vessels; see Appendix B). The results reported by Sharma et al suggest that the first case accounts for less than 2% of exams and the second case for less than 25% of exams. Therefore the absolute value of VR does not accurately reflect the proportion of exams correctly performed. However, VR is entirely adequate when used as a QC metric to compare the relative performance of sonographers given that it is consistently defined and all sonographers have the same chance of scanning a woman with a right oophorectomy or visualising a woman’s Iliac vessels, but not her right ovary.

There are more serious objections to the way ovarian cancer screening trials use VR as a QC metric for TVS scanning. First, its definition is not widely agreed, so comparisons of VR between trials can be problematic as described in Chapter 9. Second, it is subjective and suffers from significant inter-observer variability and intra-observer variability as reported by Timmerman et al³. Third, it is almost always self-reported by the examining sonographer and so cannot be considered independent. This last issue is particularly important as the over-statement of VR by sonographers in UKCTOCS (as reported in the studies contained in Chapters 8 and 9) suggest that the results reported by Sharma et al² may have been influenced by a significant source of bias which was neither addressed nor discussed in the paper.

The penultimate sentence of the Sharma et al² paper states “*In conclusion, if there are robust and monitored QA processes, ovarian VR as a performance indicator of the quality of ultrasound screening will improve with time despite adverse factors such as aging*”. This suggests that VR could be a good QC metric, but clearly only if measured correctly. In this regard the self-reporting of VR by the examining sonographer is problematic, as indicated by the gross over-reporting reported in the

studies involving its independent assessment; see Chapters 8, 9. Therefore it may be concluded that accreditation and training seem to have resulted in improvement of TVS scanning quality, but robust independent and objective QC of the examinations would be needed to allow the impact of these measures to be quantified. The following chapters explain how this might be achieved by using software tools to analysis the images from such examinations.

Summary

This chapter has provided a context for the application of the tools described in the next six chapters in terms of their use in a QI programme. The focus has been on QI in NHS England as this is where UKCTOCS operated, so is the most likely first implementer of a national screening program for ovarian cancer. However, an essential precursor for such screening is the development of an effective QI programme for TVS scanning. This remains a daunting challenge at an organisational level for the following reasons.

1. Translating techniques from one industry to another is often difficult and in this respect QI methods like Six-Sigma are particularly challenging as they typically require a change of cultural within organisations that seek to adopt them. For organisations involved in the fabrication of expensive parts to high tolerances, it is less difficult to undertake such change as the benefits are easy to envisage and the implementation is readily copied from a variety of sources. However, for healthcare organisations the gains are much more difficult to quantify and there is a lack of case-studies from which they can learn. This means the changes involve substantially more risk and need much greater resources. NHS England has seemingly little appetite for such risk and faces huge competing demands on its resources.
2. A significant barrier to the wide-scale adoption of QI in NHS England is its lack of capability in implementing the required methodologies. This is because after a series of high profile scandals involving the delivery of substandard care the quality efforts of NHS England are currently focussed on preventing failure. Accordingly its focus lies with QA rather than delivering continuous quality improvement. . Therefore, it can be argued that NHS England is not yet ready to embrace a methodology like Six-Sigma.
3. NHS England has undergone a series of reorganisations which have impeded the formation of a unified national QI strategy. It remains to be seen if the recently reformed NQB can address this problem, but in the meantime the implementation of QI remains fragmented with policy being formed at the level of various professional bodies. This means that an NHS worker's view of QI depends upon her professional connections. In comparison at a

company like Toyota everyone from the CEO in Aichi Japan to the apprentice widget maker in Deeside Wales share common values about quality and follow the same basic practices; an approach that has allowed Toyota to become one of the world's biggest and best car-makers¹³⁷.

It seems unlikely that any unified model for QI will emerge in NHS England within the next five years even though, as observed by Nicolay et al⁶⁶, healthcare systems are failing to deliver the same levels of quality as systems in other industries.

Therefore TVS screening for ovarian cancer may take a localised approach to QI similar to that used by the Fetal Anomaly Screening Programme (FASP). However, as discussed in Chapter 10, such an approach does not preclude taking full advantage of the tools and practices from methodologies like Six-Sigma, nor does it prevent the building of culture for people involved in this type of screening that puts quality at its centre.

The reports from large-scale clinical trials using TVS scanning for ovarian cancer screening have highlighted some of the difficulties that arise when performing QA on this type of imaging modality. There is clear advantage in the approach of collecting a range of QC metrics to detect scanning centres or sonographers performing in a non-standard way in terms of producing proportions out of the normal range for test-positive, test-negatives, ovary visualisation, mean ovary volume or similar measures. A key part of such a QA plan involves the confirmation of VR metrics by the practices of randomly sampling TVS exams and independently reviewing the static images or repeating an examination during the same patient visit. However, it must be recognised that the quality of TVS scanning is ultimately dependent on the skill and experience of the sonographers and in this respect both UKCTOCS and PLCO seem to have succeeded in creating effective processes for recruitment and training of such staff, though their methods differed. The UKCTOCS innovation of creating a programme of annual accreditation for their sonographers combined with targeted training and support seems to have been particularly successful in terms of improving TVS quality even though it could not be reliably quantified.

It seems probable that the difference between the PLCO and UKCTOCS approach to QA is explained in part by their funding. PLCO was clearly well resourced which

permitted a more rigorous approach to QA in terms of both staffing and IT support. Any future large-scale use of TVS scanning for ovarian cancer screening needs to be similarly well funded, particularly in respect of IT provision because the effectiveness of any QA programme ultimately depends on the quality of its QC data. In this regard the following sections of the thesis may help find solutions in terms of the development of tools to provide an objective and independent way of collecting QC metrics for TVS scanning (Part B) as well reports of their use in research studies (Part C).

Tool Chapters

Chapter 4: Ultrasound Record Archive and Associated Tools

Introduction

The static images gathered from a TVS exam form an important record, particularly those images used to measure the ovaries and endometrium. UKCTOCS decided to create a computer system which became known as the Ultrasound Record Archive (URA) to store the images from their TVS exams. It was anticipated that such an archive would allow on-going quality control of the exams as well as providing a unique resource for future research. This chapter describes the requirements for the archive, together with details of its design and implementation. The uses of the archive are also described as well as plans for future work.

Note: The URA was developed by the author as a project for his MSc studies at Cranfield University in 2007. However, the following material either cites from the MSc thesis as indicated in the text, or is new work derived from this field work in accordance with UCL regulations.

History of URA

When TVS screening started in June 2001 a process was setup to transport the TVS static images from the 13 trial centres to the central coordinating centre at Maple House, London. This process required the trial centres to copy the images stored on their Ultrasound machines to Magneto-Optical (MO) disks and send them by weekly courier to Maple House. At the outset no plans were made for the processing of the disks so they were just copied to Compact Disks (CDs) which were stored in a cupboard.

A small project was initiated in February 2005 to process the growing collection of CDs generated by UKCTOCS. The project was undertaken by the author as part of an MSc Bioinformatics course. It started by searching for existing commercial off-the-shelf (COTS) products. However, budget limitations precluded many products that might have been adopted. For example, as reported in the MSc thesis, Medicor Imaging quoted £0.40 per image which translated to £560K for the 1.4 million images it was thought would be generated over the course of the trial. A search was also made for open-source solutions that might be modified and used by UKCTOCS. However, no suitable systems were found. Therefore the decision was made to develop a bespoke system which became called the Ultrasound Record Archive (URA). The budget for the project was approximately £2,000, but was considered viable as it was anticipated that the development work would be done without charge.

The URA v1.1.0.8 passed its user acceptance testing (UAT) and was successfully released in August 2007. It has two main components; the database server and client application for importing and exporting images. It was supplied with an installation program and user manual. The URA was eventually released as the OSPACS¹³⁸ open-source project, but has not attracted an active developer community despite being downloaded more than 10,000 times from 2007 to 2015. This is probably due to the difficulty of creating a general solution that can be customized to satisfy individual requirements. Some suggestions for addressing this issue are made at the end of the chapter.

Requirements Analysis

The requirements for the UKCTOCS image archive system were stated in the MSc thesis and are summarized in the following sections.

Integration with the UKCTOCS ultrasound scanning workflow

Images from the ultrasound machine were copied onto an MO disk at the trial centres. These disks were then sent by courier to the coordinating centre at Maple House in London for transfer into the archive system.

Processing disks from trial centres

The disk format and file structure was compatible with the Microsoft MS-DOS¹³⁹ operating system. A top level directory was created for each month and individual subdirectories were created for each day during the month with a further set of subdirectories created for the images from a particular TVS exam performed during that day. The archiving system needed to process each disk to find all the exams it contained as well all the images for each exam.

Validation against data in TMS

The images for each volunteer exam were stored as files in DICOM format, as described in Appendix H. The archiving system was required to obtain the volunteer name, Volunteer Reference Number (VRN) and exam date from specific text fields in the header of the file. This information was then cross-checked with the Trial Management System (TMS) to confirm that a TVS scan record had been created for the given volunteer on the given day and that the spelling of the volunteer name in the header matched the spelling recorded in the TMS and the volunteer's reference also matched.

Manual correction

The archiving system operator was required to manually intervene when the information in the DICOM header did not match the information in the TMS. For example, the operator would be presented with a dialog box in order to correct an evident error in the name if the header contained a VRN value of 123456 and the volunteer name 'Scott', but the TMS had returned the name 'Stott' for this reference.

Similarly, errors made by the Sonographer when entering the volunteer reference were addressed by the system displaying a list of volunteer references for the given name, so allowing the correction of numeric transposition errors. More complex typographic errors involving mistakes in name, volunteer reference and/or date would require more detailed investigation, but these were considered rare.

Anonymization

In order to make the archive more useful for research, the images needed to be anonymized before storage. This required the name to be removed from the DICOM header as well as from the top corner of the image. In this way the archive would only contain images that could be identified by the volunteer reference. An advanced form of such anonymization is described in Appendix F.

Image Storage

Once anonymized the images needed to be stored in the archive using a unique numeric reference generated from the exam date and volunteer reference. This reference was required to conform with DICOM standards – i.e. it had to be a properly formed DicomRef. The volunteer reference and exam date were also required to be stored with each image record.

Image Export

An export facility was needed to copy a collection of images corresponding to a list of DICOM references to a specified folder on a computer. It was anticipated that this list would be obtained by querying the URA database for records matching a list of volunteer references and exam dates as generated by the Trial Management System (TMS).

Non-functional Requirements

The non-functional requirements included an installation program and on-line help manual. It was subjected to unit and functional testing before being deployed into the production environment for final user acceptance testing as described in the MSc Thesis.

Design and Implementation

The URA was implemented as a thick client-server solution integrated with the other components of the UKCTOCS ultrasound data processing system as shown in Figure 4-1. The servers were located in Maple House within an office used as a machine room.

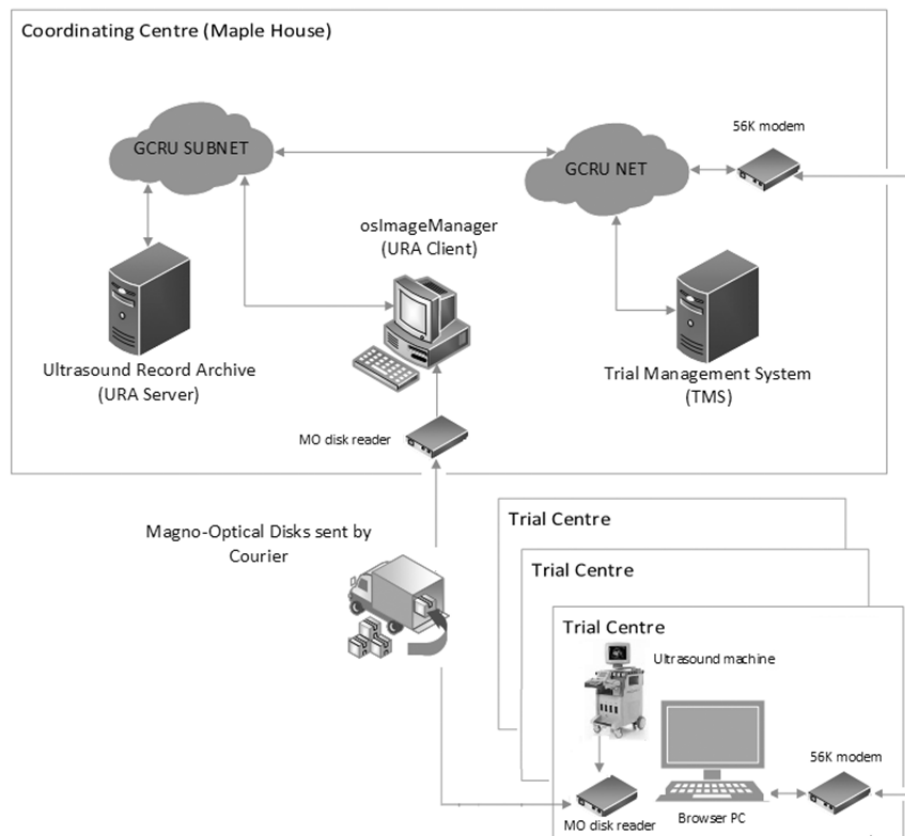


Figure 4-1: UKCTOCS Ultrasound Data Processing System. After the TVS examination information from the Scan Report Form (Appendix C) is entered into the TMS using a web form presented on the Trial Centre Browser PC. The static images from the examination were initially stored on the ultrasound machine, but later copied onto a MO disk and couriered to the Coordinating Centre for entry into the URA using the osImageManager application.

The URA server was built using a Dell Power Edge PC running Windows Server 2000 (SP2) and SQL Server 2005. Each image was stored in the 'DicomFile' field of a record in the 'Image' table of the 'osImage' database created on the URA Server. The datatype of the DicomFile field was 'image' so that the entire DICOM file was serialised as binary large object data. Other fields in the Image table record included 'CaptureDate' (date of the TVS exam as stored in the DICOM header), VolunteerRef

(UKCTOCS identifier for the person being scanned), DicomRef (unique identifier for image), DiskRef (unique identifier for the MO disk).

The interface between the server and the thick client was implemented by stored procedures on the server side and .NET libraries using standard ODBC drivers on the client side. Access to the stored procedures was granted by user group to facilitate the implementation of system security. The thick client was called osImageManager. It was developed in C# using Visual Studio 2005 and operated on a standard Desktop PC running Windows XP (SP2). The ezDICOM¹⁴⁰ (rev22) ActiveX control was used to display the images in the main window as shown on the right hand side of Figure 4-2.

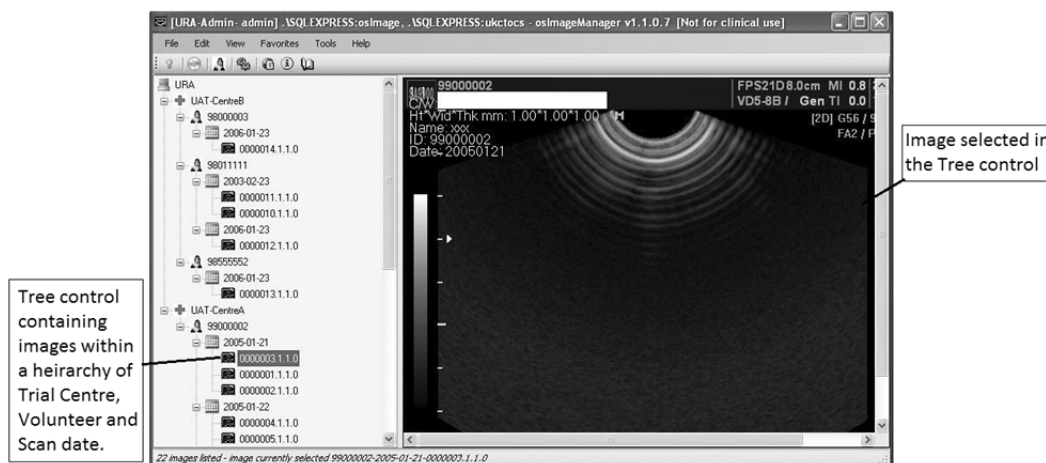


Figure 4-2: osImageManager Main Window. Images imported into the URA can be displayed in the tree control in the left hand side of the main window. This allows them to be displayed in a hierarchy with the URA forming the root and a succession of nodes for Trial Centre, Volunteer and Scan Date. The Scan Date node corresponds to a particular TVS examination performed on the volunteer on the given scan data and contains a collection of nodes for the images recorded by the sonographer during scanning. Whenever the user clicks on the image node its corresponding image is displayed in the right side of the main window.

Discussion

The URA was used to import images from TVS examinations from August 2007 until after the scanning was completed at the end of 2011. The archive was also used for quality control and research purposes. These applications of the system are described in the following sections.

Use of osImageManager for importing Trial Images

The primary function of the URA osImageManager application was to allow operators to import images from TVS exams into the osImage database from the CDs in which they had been temporarily stored. This required the operator to insert the disk into the reader attached to the client PC and then open the File Import dialog; see Figure 4-3. After importing the disk its contents were displayed in the main window of osImageManager (see Figure 4-2) to provide confirmation that the operation had succeeded.

The imported images were displayed in the main window tree control in a hierarchy of Trial Centre, Volunteer and Scan Date. Typically a disk would contain images from just one Trial Centre and each Volunteer would have just a single Scan Date. The number of images for each scan ranged between 1 and 31. Clicking an image reference in the tree control would result in the corresponding image being displayed in the right hand side of the window.

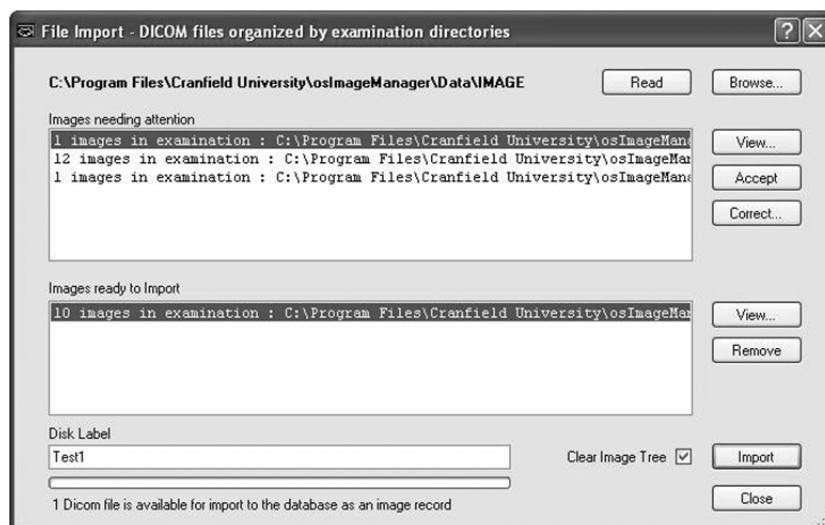


Figure 4-3: URA File Import Dialog. When the user clicks the 'Read' button the contents of the disk are read and analysed. Examinations with images that are ready for importing are displayed in the bottom list whilst the top list contains examinations with images that need some form of correction. For example, the volunteer reference may not match any record in the URA due to a typographical error. In such a case the user selects the exam and then clicks the 'Correct' button in order to open a dialog box that allows the problem to be fixed. Afterwards the user clicks the 'Accept' button to move the exam to the bottom list. After all the problems have been resolved the user clicks the 'Import' button to copy the images for all the exams in the bottom list into the URA database.

Processing and Storage of Images from Trial Centres

UKCTOCS employed a data entry clerk to process the MO-Disks arriving at Maple House. The time required to process each disk varied significantly, with some disks being processed in less than 10 minutes and others taking several hours. This variation resulted from the number of images recorded on the disk that required some form of manual intervention – i.e. correcting typographical errors in the volunteer name, reference or scan date. It was observed that some trials centres generated more errors than others in these respects. Though there was no attempt to systematically report error rates and provide feedback in order to improve the overall process, some problems were discussed with individual centres informally.

The data entry clerk required a few hours initial training before gaining the proficiency needed to import the images without supervision. The system proved itself reliable and easy to use. However, finding the information needed to correct typographical errors proved difficult in a small number of exams, particularly in the case of typographic errors in both name and volunteer reference. In addition, processing throughput was suboptimal due to the need to correct the same errors for each individual image in the exam once the right information had been found.

MO-Disks arriving at Maple House were processed and entered into the URA only spasmodically. Therefore, in cases of exams that had reported abnormalities or other issues, the corresponding disks would be inspected by the National Lead Sonographer soon after arrival. In such instances the images would be reviewed directly from the disk before being returned to the queue for subsequent URA processing. It was observed that approximately an hour was needed to locate and review the images for each of these exams. Therefore quality control was not performed on images for exams unless an exception had been detected in the Sonographer's report. The use of the URA for quality control during the trial was limited to the review of seven sonographers' work by nine experts in 2009 (see Chapter 8) and the review of 1,000 TVS examinations by one expert in 2015 (see Chapter 9).

The backlog of MO-Disks that had accumulated at Maple House was reduced by a series of initiatives involving the employment of a contract data clerk to work continuously on the task over a number of months.

Data Stored in URA

Details of the number of TVS Exams stored in the URA in 22 July 2015 are given in Figure 4-4. The distribution of these exams by trial centre is shown in Figure 4-5

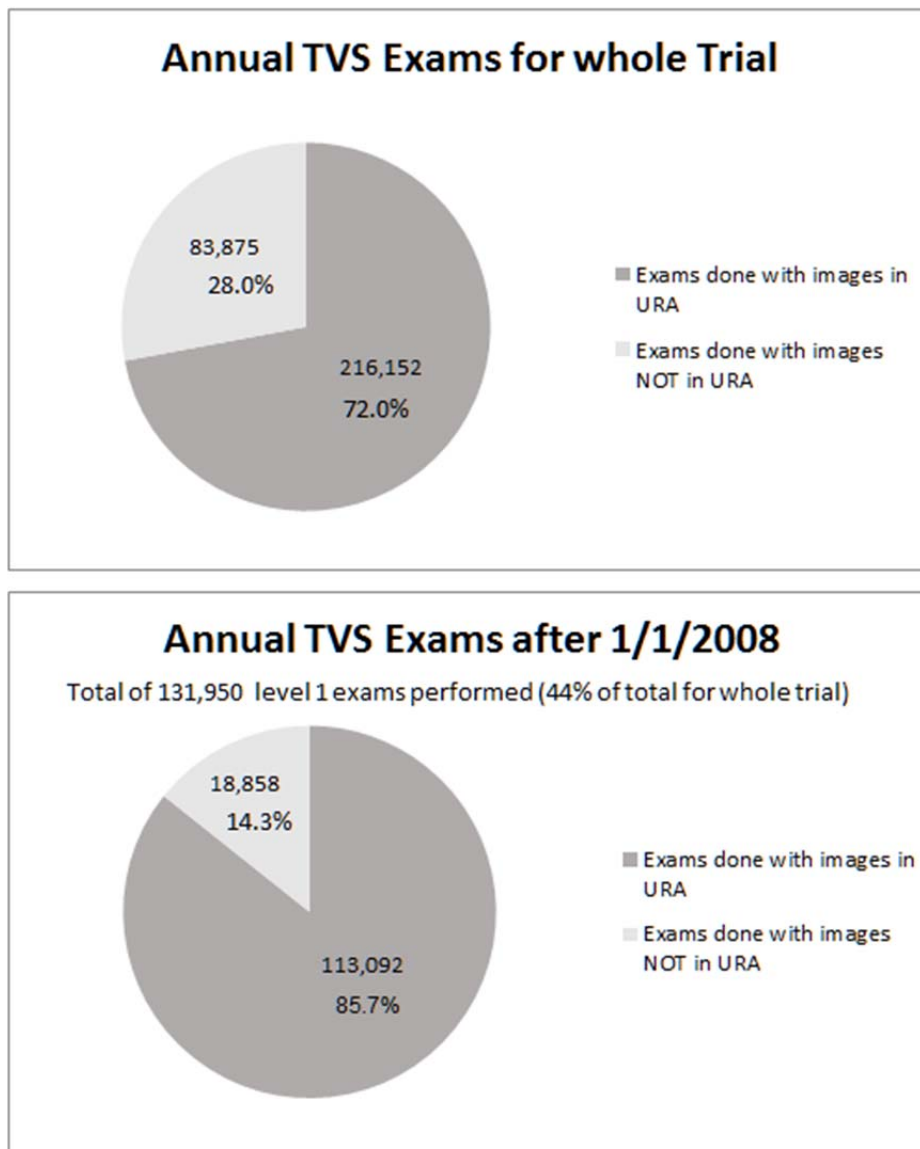


Figure 4-4: UKCTOCS TVS Exams stored in the URA with respect to total exams performed. The left side shows that 72% of all the exams performed during the trial have been stored in the URA. The right side shows that this figure increased to 85.7% after 1st January 2008 when the machines were changed. Appendix A shows the source of these images.

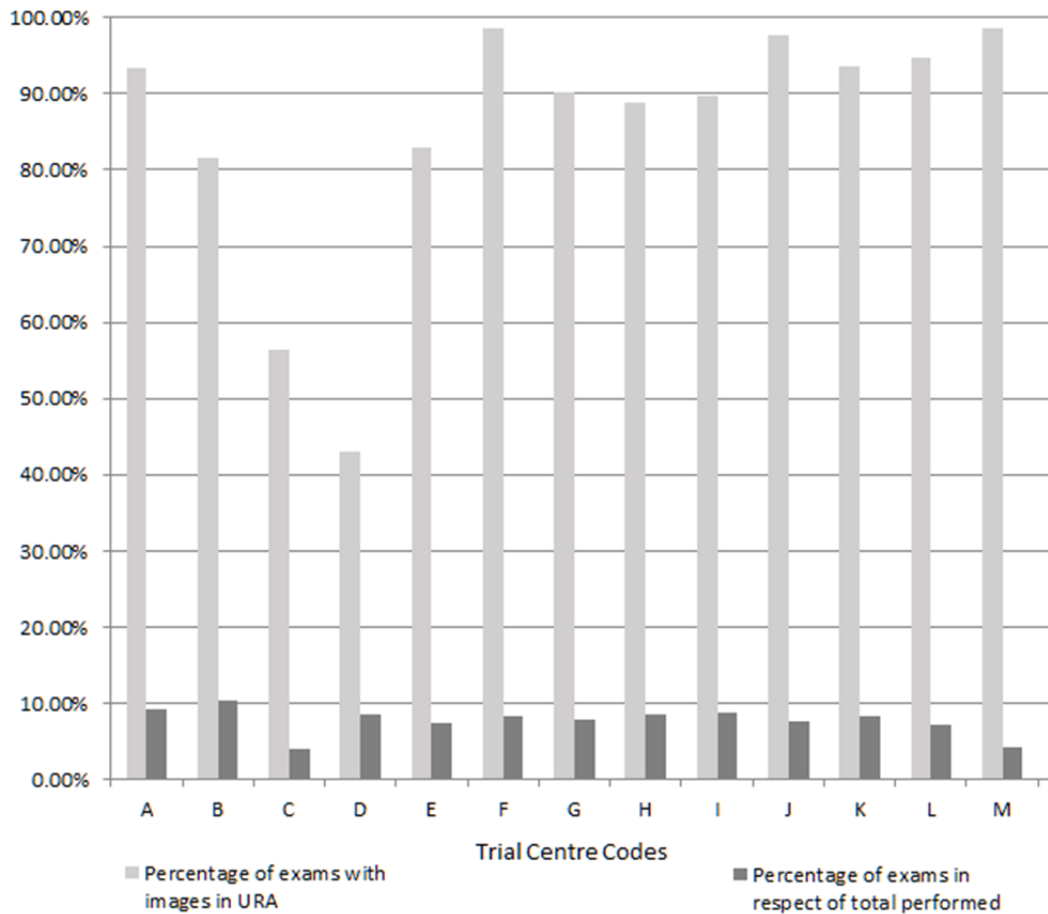


Figure 4-5: Distribution of TVS exams performed after 1st January 2008 is shown by Trial Centre. The light grey bars show that most of the thirteen trial centres provided images from at least 80% of the TVS exams they performed, though centres C and D provided a much lower proportion. The dark grey bars show that all centres performed a similar number of exams except for centres C and M which did significantly less.

Use of URA for Research - osImageManager

A basic image export facility was provided by the osImageManager application to facilitate the use of the URA for research. This facility required the user to populate the tree control by running a stored procedure to list all records in the osImage database for a selected Volunteer Reference (or Disk Ref). The images in the tree control could then be exported to a selected Folder using the File Export Dialog; see Figure 4-6. The export feature allowed the creation of small datasets of images such as might be required for reviewing images from all scans performed on a particular volunteer.

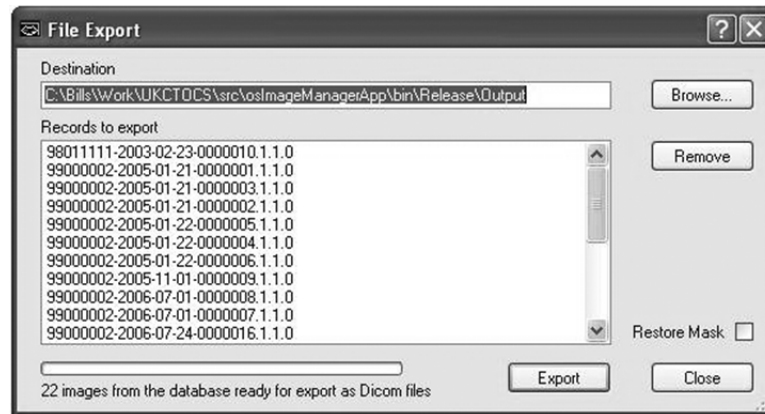


Figure 4-6: URA File Export Dialog. Images for the exams listed in the main window of osImageManager (Figure 4-2) can be exported to the folder specified in the text box at the top of the dialog.

Further Development - URAImageExport

The osImageManager application was not designed for the purpose of creating large datasets of images for research studies such as those described in Chapters 7 and 9. Therefore in 2012 a separate application called, URAImageExport, was developed to satisfy this need. It is described below in terms of its requirements and basic design. This tool was used to export the images associated with a given selection of TVS exams from the URA database to a local file system. In addition to creating the collection of bitmap files needed to contain such images, URAImageExport also created a custom XML file called an IDF (Image Dataset File) which served to list them and their properties. In this way the image datasets used for research were specified in a way that allowed them to be processed by the tools described in Chapters 5 and 6.

Requirements Analysis

The following requirements were identified for URAImageExport:

1. Read a list of image references from a Comma Separated Variable (CSV) file prepared by running queries on the Trial Management System (TMS).
2. Validate image references have corresponding records in the URA database
3. Iterate through the list of image references and for each read the corresponding image record so its serialised image data field can be converted to a BMP file and written into a specified directory.

4. Create an IDF file listing the image files that have been copied from the database to a local folder on a PC.
5. Add to the IDF file properties associated with each image as given in CSV files exported from the TMS.

Design and Implementation

URAIImageExport was developed as a prototype. Accordingly the type of defined process used for the earlier development of the Ultrasound Record Archive was not followed. Instead the development followed a more informal approach as appropriate for a tool developed by a single person for his own use. Accordingly, it did not have a formal release cycle and lacks artefacts such as a user manual and other documentation that were provided for the Ultrasound Record Archive. Work is ongoing to develop a tool with similar functionality that can be integrated with the workflow described in Chapter 6.

Applications for URAImageExport

The process of creating an image dataset (IDF) starts with a list of image records with exam dates and volunteer references. In the case of UKCTOCS, this is typically achieved by running a query on the TMS in order to find volunteers and exam dates that satisfy the inclusion criteria of the study. The result is saved as a CSV file which can then be opened with Excel. The XML mapping facility of Excel allows the creation of a XML file with the required IDF structure as shown in Figure 4-7.

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<ImportFile>
  <Image ImageRef="1.2.826.999999.1.9999.1.20090131114342998.1.4.0" VolRef="23007432" ExamDate="2009-01-31"/>
  <Image ImageRef="1.2.826.999999.1.9999.1.20090131114342998.1.1.0" VolRef="23007432" ExamDate="2009-01-31"/>
  <Image ImageRef="1.2.826.999999.1.9999.1.20090131114342998.1.2.0" VolRef="23007432" ExamDate="2009-01-31"/>
  <Image ImageRef="1.2.826.999999.1.9999.1.20090108140416840.1.1.0" VolRef="11070410" ExamDate="2009-01-08"/>
</ImportFile>
```

Figure 4-7: Format of URAImageExport input file. The file lists the images that need to be exported from the URA database. ImageRef corresponds to a unique key in the image record, but VolRef and ExamDate are also provided to allow a consistency check. The format of the IDF file has evolved since the creation of URAImageExport so the ICDatasetApp program (described in Chapter 5) allows translation from a number of formats.

URAIImageExport populates its input list shown at the top of Figure 4-8 with the contents of the IDF file which is opened using the menu command ‘File | Open’. The user then maps the names of the fields in the file to the names of the fields in the database before clicking ‘Validate’ to check that the corresponding records are

present and correct. Any errors are displayed in the messages list and validated records are displayed in the output list shown at the bottom of Figure 4-8. The user can then export the images to a local folder on the PC using the export dialog box.

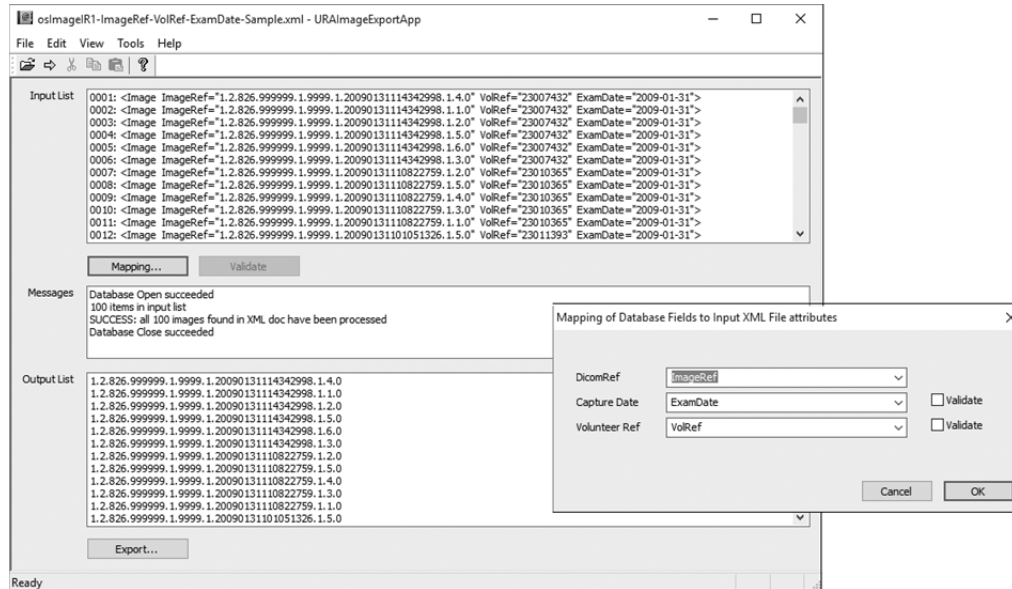


Figure 4-8: URImageExport main window. The fields in the URImageExport input file (Figure 4-7) are mapped to corresponding fields in the URA database. This feature allows URImageExport to operate with other image databases.

URImageExport was developed with the primary aim of supporting the osImage database implemented by the URA, but it also provided an opportunity to build a tool that might support image databases with different schema. Therefore the SQL query used to produce the image dataset could be edited, though security restrictions were applied. It was also possible to define the names of the fields in the XML record which would be mapped by the application to the image reference, date of examination, and personal identifier (Volunteer reference in the case of UKCTOCS) in the database.

Future Plans

In 2016 the URA and TMS servers will be moved from Maple House to a dedicated UCL datacenter. This will improve data security in terms of resilience to external threats as well as hardware failure. It should also allow definition of quality of service and provide higher bandwidth for data access. In the longer term,

consideration should be given to creating a more general biobank able to service the data/sample storage and processing requirements of UKCTOCS as well as other studies. This work may take advantage of the eMedLab infrastructure being constructed by a consortium comprising UCL, Queen Mary University of London, London School of Hygiene & Tropical Medicine, the Francis Crick Institute, the Wellcome Trust Sanger Institute and the European Bioinformatics Institute; see Appendix J

Proposed Institute for Women's Health Biobank

The creation of general biobank would allow the Institute for Women's Health to build upon its reputation for running clinical trials by offering an established IT infrastructure which could be configured at low cost to meet the needs of a wide variety of trials. It would also encourage collaboration with other organizations by providing a standard platform for sharing data and its analysis, again at low cost.

The architecture of an IfWH Biobank was proposed in 2013¹⁴¹ based on the needs of UKCTOCS, but with consideration given to its use for storing artifacts generated by other clinical trials. This architecture envisaged a number of interfaces to allow interoperability between different components developed either for general use or to satisfy the needs of particular trials. The main components of the system are shown in Figure 4-9.

Biobank Management System (BMS)

The Biobank Management System is responsible for user/data security and reporting as well as trial and research study workflows. It interfaces with other (sub) systems such as the Image and Trial Data Management Systems, which are specific to particular trials such as UKCTOCS. These interfaces need careful consideration in order to minimize the amount of work required to add a new trial to the Biobank. Consideration also needs to be given to the trade-off between the significant amount of work necessary to create general purpose sub-systems which can be re-used by a number of different trials and the creation of sub-systems specific to a particular trial which may individually require much less effort to create, but might collectively represent much greater effort.

Study Data Management System (SDMS)

The Study Data Management System would be responsible for enhancing the content of the Biobank by adding information about its samples and data. For example, in respect of a dataset of ovarian ultrasound images reviewed by experts in gynaecological scanning, value would arise from recording the results in the SDMS; ovary seen, errors in calliper placement, etc. Successive studies might add further properties so aggregating information about certain images in the Biobank. This would enhance the value of the Biobank by allowing datasets to be prepared on the basis of these additional properties.

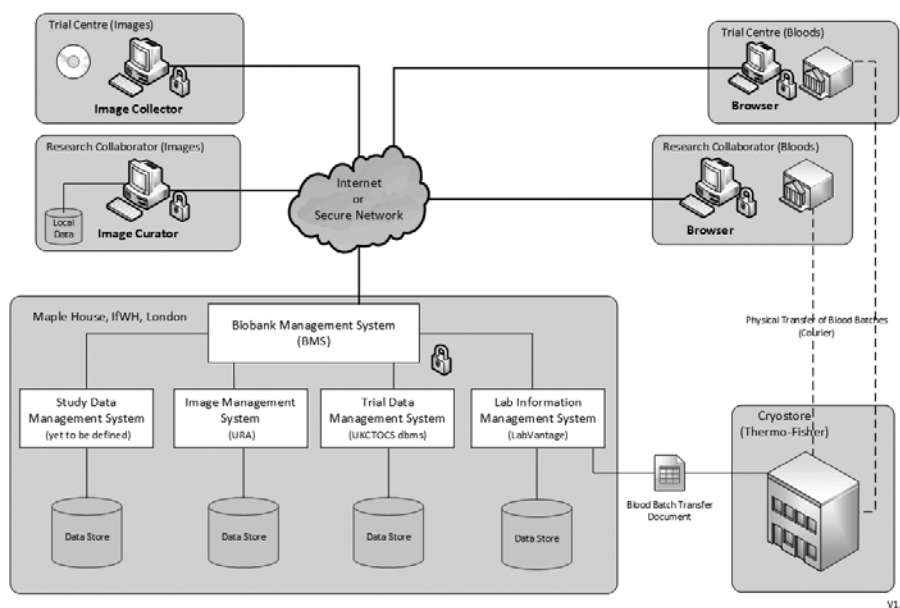


Figure 4-9: Proposed Architecture for Institute for Women's Health Biobank. The development of a general purpose BMS with common interfaces for a series of adapter modules could allow integration of the various systems used by UKCTOCS as well as other trials.

Image Management System (IMS)

The Ultrasound Record Archive (URA) is a form of Image Management System (IMS), but is specific to UKCTOCS. It could be replaced by a more generic IMS providing functionality that would support images captured by a wider variety of trials:

- Importing images in various formats from a variety of sources instead of just reading DICOM files from a specified directory structure on an MO-Disk.

For example, the implementation of the Picture Archiving Communication System (PACS) protocol allowing importation of images, *via* a data communication network, directly from an imaging machine or Image Collector application (as described below).

- Support for the sort of generalised image processing described in Chapter 6 in order to prepare images for storage. For example, a domain specific language could be used to anonymize images by specifying the areas to cut and fields in the header to remove. This could implement the type of enhanced anonymization and GUID identifiers described in Appendix F.
- Storage of the images in a variety of structures with suitable references to the examination data stored in the TMS. For example, a collection of images could be associated with a particular exam like in UKCTOCS. Alternatively, a sequence of images could be stored to form a video, tomographic sequence, etc.
- Integration with the TMS to allow the listing of images matching specified criteria. For example, all images for a given examination or examinations meeting criteria such as belonging to a particular volunteer.
- Export of a dataset of images from a list of references, such as those matching specified criteria. This might be used to create a collection of images for research purposes or for use in a periodic quality control review. The generation of an Image Dataset File (IDF) for this image collection would allow integration with the image processing tools described in Chapter 6.

Trial Management System (TMS)

The UKCTOCS database management system (DBMS) is a SQL Server solution which implements the Trial Management System (TMS) for UKCTOCS; see Figure 1-5. It contains information such as the entire medical history of the volunteers recruited for the trial, the Sonographer report of each Ultrasound exam and the data gathered from the CA125 (Blood) arm of the trial.

It would be possible to design a TMS able to support multiple clinical trials by separating information that is common to all trials from that which is specific to an individual trial. In this way the DBMS schema could be adapted for the needs of new

trials whilst retaining support for existing trials and providing a subset of functionality common to all. The success of such a design can be measured by the amount of work needed to support a new trial, as a good design will result in this effort being much less than would be required to implement a separate TMS specific to the new trial.

Historically, changing the schema of a SQL database used in a production environment was considered challenging. However, the effort required to implement such alterations in a modern system has been significantly reduced with the introduction of change support tools¹⁴² (SQL Unit Test, SQL Source Control, SQL Compare) and database refactoring techniques¹⁴³. The problems of changing a TMS schema might also be addressed by implementing a document-oriented database such as MongoDB. Such databases feature dynamic schema, which can be changed with greater control and fewer implications for data resynchronisation.

Lab Information Management System (LIMS)

A collection of spreadsheets was initially used by UKCTOCS to track blood samples removed from the cryostore. There were plans in 2014 to replace this adhoc system with LabVantage¹⁴⁴, a commercial LIMS. Like most commercial LIMS, LabVantage enables integration with other systems by providing interfaces implemented with standard technology such as Simple Object Access Protocol (SOAP) based Web services. In this way the LIMS could be integrated with other components of the proposed Biobank. However, the migration to LabVantage was not completed due to cost constraints. Therefore at present blood sample tracking continues to be managed using spreadsheets.

Application: Image Collector

It was proposed that an application called Image Collector should serve to validate and then transfer information to the Biobank IMS and TMS. Implementing such an application as a thick client or web-based application at the trial centres would allow validation to be performed at the time of the examination, rather than months later; the approach adopted by UKCTOCS. Performing validation at examination time allows errors to be corrected more efficiently and directly uploading images to the IMS avoids the significant cost incurred by UKCTOCS of importing them from MO-

Disks. It would also be possible to make Image Collector responsible for anonymizing the images (see Appendix F) so they could be transferred to the IMS over a public network, though a secure data connection would still be required.

Application: Image Curator

The Image Curator application is described in Chapter 5, both in terms of its current implement and future plans. Its primary function is the curation of image datasets in order to add information that increases the value and utility of the images. For example, a group of experts might review an image dataset and score them in terms of quality and presence of objects like ovaries, complex cysts and other abnormalities. Such curation adds value in respect of research and quality control.

Web Browser: Trial Centre Collaborator

Information collected by a Trial Centre Collaborator during the running of a trial could be submitted through a web browser and website integrated with the BMS. This is the same approach adopted by UKCTOCS for the submission of Sonographer and CA125 reports. However, better accuracy and greater efficiency would be achieved by improving the workflow and integration with trial machines; see Chapter 10. Consideration also needs to be given to validating such information as well as maintaining its security.

Web Browser: Research Collaborator

Information about a clinical trial should be made available to a Research Collaborator through a web browser and website integrated with the BMS. Consideration would need to be given to controlling the scope of such information given to Research Collaborators as well as regulating security, intellectual property rights, data protection, ethical use and so forth. It is likely that different trials hosted by the proposed Biobank would have different requirements in these regards, so different web applications may need to be developed. However, it is expected that the effort would be less than would be necessary when developing an application without utilising existing Biobank infrastructure.

Proposed open-source project

A number of open-source projects have been created to support laboratories and clinical trials; see examples listed in Table 4-1. Some of these projects target applications involving biological sample collection and others target electronic data capture (EDC). Support for other artifacts, particularly images, is lacking. Therefore it is proposed that the open-source OSPACS project should undertake a redesign of the URA in order that a Biobank of the type outlined in Figure 4-9 could be built by integrating it with one or more open-source products of the kind described in Table 4-1.

Name	Comments
Bika LIMS ¹⁴⁵	LIMS for Bioinformatics includes sample management and data distribution. It is supported by a commercial organization selling its services.
OpenSpecimen ¹⁴⁶	Specimen management system originating from the U.S. National Cancer Institute funded caBIG programme and now extended to support any type of disease and specimen type. It is supported by a commercial organization selling its services. Installed in 25+ institutions.
OpenELIS ¹⁴⁷	System with support for sample management and workflow. Poor documentation makes modification difficult. Supported by certified implementers.
LabKey Server ¹⁴⁸	Its strength lies in data analysis, but also provides specimen management and other LIMS functionality. Commercial support is provided, though it is expensive relative to other systems.
Open-LIMS ¹⁴⁹	LIMS for Bioinformatics with sample management and workflow facilities. However, it doesn't seem to have been widely adopted and doesn't have commercial support.
OpenClinica ¹⁵⁰	An electronic data capture (EDC) system for clinical research. It provides features to capture, clean and manage clinical trial data. OpenClinica has commercial support.
ClinCapture ¹⁵¹	EDC system developed by Clinovo, a commercial clinical research organization. This product aims to support the creation and management of studies. It features form generation, workflow rules, randomization, calendared events, etc.
ObiBa	Provides software solutions for epidemiological data management and analysis including Onyx (design and administration of questionnaires and EDC), Opal (data store), DataSHIELD (data analysis for Opal), Mica (reporting, query), Agate (authentication server). ObiBa is developed by Maelstrom Research which is supported by a variety of funding agencies.

Table 4-1: Examples of Open Source Biobanks. Adapters could be developed for these systems to allow integration with the Biobank Management System (BMS) shown in Figure 4-9. In this way the Institute for Women's Health might create the infrastructure needed for their clinical trials at a relatively low cost.

In addition the OSPACS project should attempt to implement support for the type of image curating and processing tools described in Chapters 5 and 6 as well as permitting image storage and processing in some form of 'cloud' solution; see Appendix G. This would significantly reduce the cost of performing clinical trials involving images as well as encouraging the development of advanced image processing. It is hoped that such an initiative would lead to innovative image diagnostics in much the same way that the advent of tools for the analysis of genomic data have facilitated advances in personalized (translational) medicine.

Summary

The URA contains an archive of images from 300,027 TVS annual exams performed by UKCTOCS on 480,250 volunteers over a ten year period. The mean number of annual scans performed on these women is 6.81 which suggests images from almost seven years of serial scanning is available for most of them. In addition all these women were flagged with the national cancer registries and follow-up medical details exist for 98.9% of women in the ultrasound arm of the trial. This makes the URA a unique resource which is unlikely to be surpassed by future work. In addition, the URA adds significant value to the trial by allowing retrospective quality control to be applied to the TVS examinations so that high quality data can be selected for the purposes of research, as suggested in Chapter 9.

The following two chapters describe some software tools that have been developed to use the images in the URA not only for the purposes of automating quality control, but also to assist future research on the images themselves.

Chapter 5: Tools for TVS Image Reviews

Introduction

The Quality Control (QC) of TVS scanning used for ovarian cancer screening needs to be considered at a number of levels, as described in Chapter 3. However, in respect of the QC of the examination itself, some form of independent manual review of the static images captured during examination is clearly desirable. Such reviews usually involve experts in gynaecological scanning viewing images from a sample of exams and recording their impressions in respect of a given set of properties. For example, the experts might be asked to set a binary variable according to whether they consider the image shows measurement of a feature that resembles an ovary. The main limitation of this approach to the QC of TVS is the significant manpower resources required to conduct the reviews. Therefore software tools able to facilitate the manual review process may help address this issue. This chapter describes three such tools developed to enable image reviews of examinations performed by UKCTOCS for QC as well as for research purposes.

The type of tool used for manual reviews reflects the number of images being considered, the type of properties that need to be assessed and the number of reviewers. Consequently the chapter starts with the description of the simple spreadsheet used by a single expert to confirm ovary visualization for 1,000 TVS Exams, as reported in Chapter 9. This is followed by the description of a bespoke software application developed specifically to help a team of nine experts review a collection of properties from 357 TVS Exams, as reported in Chapter 8. The chapter concludes with an account of another bespoke software application, but one developed to provide a general solution to the problem of performing image reviews. Although this application was not used in a UKCTOCS image review, its ability to allow experts to identify the exact boundaries of objects in images, such as ovaries, demonstrates its potential for future use in a wide variety of reviews, particular those associated with creating datasets for algorithm development.

Spreadsheet used to confirm ovary visualization

A spreadsheet is often used for performing a manual review of TVS images. Typically it contains a row for each image as well as columns for properties like a link to the associated file, a field to record the expert's opinion, etc. The requirements, design and implementation of such a system are described below together with a discussion of its use for the review.

Requirement Analysis

UKCTOCS decided to perform an expert review of images from 1,000 TVS exams with the aim of confirming that the images used to measure the ovaries had callipers marking objects that were ovaries rather than, say, bowel. Details about this review are given in Chapter 9. The tool requirements can be summarised as:

- For use by only one reviewer
- Reviewer able to review each image in turn and record his assessment in terms of a free text comment and the following categorical variable values:
 - **Not Done** – image not reviewed (initial condition at start of review)
 - **Left OK** – image contains left ovary and correctly measured
 - **Right OK** – image contains right ovary and correctly measured
 - **Left Bad Measure** – image contains left ovary, but is not correctly measured
 - **Right Bad Measure** – image contains right ovary, but is not correctly measured
 - **Left Not Visualised** – image annotated as left ovary, but callipers do not measure an ovary
 - **Right Not Visualised** – image annotated as right ovary, but callipers do not measure an ovary
 - **Left other problem** – image annotated as left ovary, but is not an acceptable for the purposes of measuring an ovary
 - **Right other problem** – image annotated as right ovary, but is not an acceptable for the purposes of measuring an ovary
 - **N/A** – image was not used for ovary measurement. For example, the image was used to measure endometrium thickness.

- The reviewer when viewing the image needs to be able to adjust brightness and contrast
- The reviewer is provided with information to assist in identification of the images used to measure the ovaries as provided by the workflow software – see Chapter 7
- It must be possible to analyze the reviewer assessment in order to identify exams that had not visualised both ovaries and produce descriptive statistics in this regard

No existing software product able to satisfy these requirements was found, spreadsheet or otherwise. Therefore given the trivial nature of the work, it was decided to construct a spreadsheet anew.

Design

Microsoft Excel¹⁵² was selected as a tool for constructing the spreadsheet as it supported import from XML so it would be easy to insert the image data from a workflow image dataset file (IDF), as defined in Chapter 4. In addition Excel provides the following function:

- `HYPERLINK(Link Location, Friendly Name)`

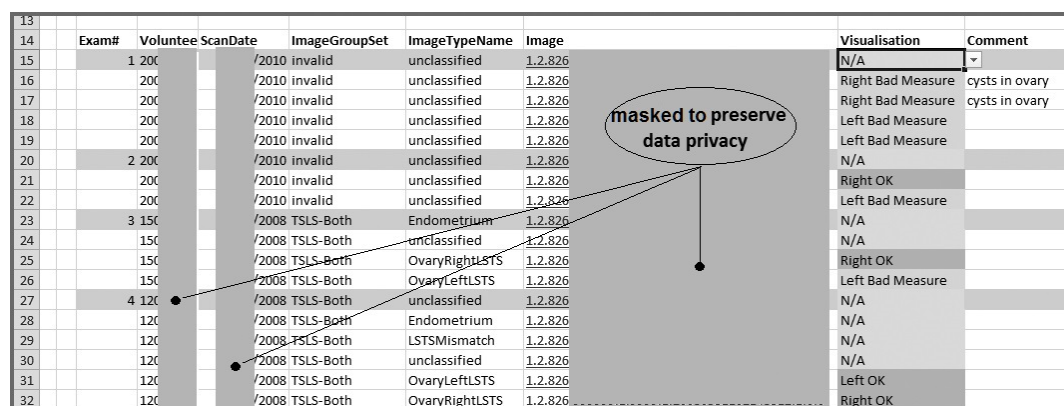
This function creates a clickable hypertext link from text in the cell defined by Friendly Name parameter using a file location given by the link location parameter. Therefore images referenced in the rows of the spreadsheet could be displayed in a graphical editor simply by clicking on the appropriate link in the spreadsheet.

The Hyperlink function opens the default editor associated with the file type of the link on the target PC. The default editor can be changed by altering values in the Windows Registry. The STDUViewerFile¹⁵³ editor was selected for displaying images as it has features allowing adjustment of contrast and brightness.

The categorical variables for the reviewer's assessment were implemented an Excel dropdown list in order to ensure that only the given values were entered as defined by the requirements.

Implementation

The dataset of images for review were exported from the URA as 640 x 480 pixel BMP files (24 bits per pixel) using the URImageExport application described in Chapter 4. These files were copied to an encrypted hard disk. This allowed them to be moved outside the security provided by the trial coordinating centre so that the review could be performed in the office of the expert in gynaecological scanning. A Personal Computer (PC) was prepared for the purposes of the review and the necessary software was installed; Microsoft Office 2010 and STDU Viewer (v1.6.375). A spreadsheet was also prepared in accordance with the design outlined above and installed on the PC; see Figure 5-1.



	Exam#	Voluntee	ScanDate	ImageGroupSet	ImageTypeName	Image	Visualisation	Comment	
13									
14									
15		1	200	/2010	invalid	unclassified	1.2.826	N/A	
16			200	/2010	invalid	unclassified	1.2.826	Right Bad Measure	cysts in ovary
17			200	/2010	invalid	unclassified	1.2.826	Right Bad Measure	cysts in ovary
18			200	/2010	invalid	unclassified	1.2.826	Left Bad Measure	
19			200	/2010	invalid	unclassified	1.2.826	Left Bad Measure	
20		2	200	/2010	invalid	unclassified	1.2.826	N/A	
21			200	/2010	invalid	unclassified	1.2.826	Right OK	
22			200	/2010	invalid	unclassified	1.2.826	Left Bad Measure	
23		3	150	/2008	TSLs-Both	Endometrium	1.2.826	N/A	
24			150	/2008	TSLs-Both	unclassified	1.2.826	N/A	
25			150	/2008	TSLs-Both	OvaryRightLSTS	1.2.826	Right OK	
26			150	/2008	TSLs-Both	OvaryLeftLSTS	1.2.826	Left Bad Measure	
27		4	120	/2008	TSLs-Both	unclassified	1.2.826	N/A	
28			120	/2008	TSLs-Both	Endometrium	1.2.826	N/A	
29			120	/2008	TSLs-Both	LSTSMismatch	1.2.826	N/A	
30			120	/2008	TSLs-Both	unclassified	1.2.826	N/A	
31			120	/2008	TSLs-Both	OvaryLeftLSTS	1.2.826	Left OK	
32			120	/2008	TSLs-Both	OvaryRightLSTS	1.2.826	Right OK	

Figure 5-1: Spreadsheet used for Image Review. The reviewer clicks on the link in the Image column to display the image and then records his or her assessment of visualization in the following column.

The system was demonstrated to the reviewer and a number of minor modifications were made to the spreadsheet as a result.

Discussion

The reviewer did not report any problems in the use of the spreadsheet and it allowed the review to be completed satisfactorily. However, it was agreed that the use of a spreadsheet would have been problematic had more than one reviewed been involved, or had the review required the gathering of more detailed information.

The review was performed in the first quarter of 2015 and took approximately two months to complete. The results were analyzed as described in Chapter 9.

Application used to conduct an audit of TVS Exams

The audit of TVS exams involving a review of the associated images by a group of experts in gynaecological scanning may require a more complex solution than a spreadsheet, particularly if there are multiple issues to review. In the case of the TVS audit described in Chapter 8 a custom application was developed specifically for the image review. The requirements, design and implementation of such a system are described below together with a discussion of its use for the review.

Requirement Analysis

UKCTOCS decided to engage a team of experts in gynaecological scanning to perform an expert review of images from a selection of TVS exams performed by the small number of sonographers who had reported ovary visualization rates of 89% and above. Details about this review are given in Chapter 8. The tool requirements were obtained after detailed user story analysis with the UKCTOCS data manager and the National Lead Sonographer (NLS) and can be summarised as:

- For use by multiple reviewers, though not simultaneously.
- Reviewer's assessments needed to be recorded securely in a way that facilitated its subsequent analysis
- The system should have a user interface that facilitates the execution of the review. Particular emphasis was given to allowing reviewers to work quickly, track their progress and easily identify any exams they had inadvertently skipped.
- Data entered by the reviewer needs to be validated so that either all the information is stored or nothing is stored.
- Each exam needed to be reviewed in two steps. The first step involved a review of all its images. The second step involved an overall review of the exam.
- Step 1 – reviewer able to record assessment of each image in terms of:
 - Setting of a categorical variable reflecting whether annotation was not given by the sonographer (unlabelled), or whether it allowed the image to be identified as a left ovary, right ovary, other feature (e.g. endometrium).

- Setting of a categorical variable reflecting whether the callipers identify an object that is a) an ovary, b) endometrium, c) inclusion cyst, d) simple cyst e) fibroid, f) bowel, g) muscle, h) vessel, i) other – i.e. none of the above.
- Setting of a binary variable reflecting whether the callipers were correctly placed to allow measurement of the ovary, or not.
- Setting of a discrete variable for image quality in the range 1 (bad) to 5 (excellent).
- Step 2 – reviewer able to record an overview of the examination in terms of:
 - Identification of the images used by the sonographer to measure the left and right ovary (see notes 1, 2)
 - Setting of a categorical variable reflecting classification of images used to measure the left and right ovary (see note 2) in terms of a) an ovary, b) not an ovary, unconvinced – i.e. reviewer is not certain, c) not imaged – i.e. the image was not used to measure an ovary. For each of these classifications (except not imaged) further data was recorded in respect of both ovaries:
 - **Ovary:** information aggregated from the images used for measurement were assessed in terms of:
 - Setting of a categorical variable reflecting whether annotation was not given by the sonographer (unlabelled), or whether it allowed the image to be identified as a left ovary or right ovary.
 - Setting of a categorical variable reflecting calliper placement in respect of:
 - Correctly placed for measurement
 - D1 and D2 not perpendicular to each other, as described in Chapter 1.
 - One or more of the measurements D1, D2, D3 were missing.
 - D3 was measured in an image of an ovary in longitudinal section rather than transverse section.

- Callipers incorrectly placed on the extents of the ovary boundaries for any of D1, D2, D3.
 - No measurement made for any of D1, D2, D3.
- Setting of a discrete variable for image quality in the range 1 (bad) to 5 (excellent).
- **Not an ovary:** information aggregated from the images used for measurement were assessed in terms of:
 - Setting of a categorical variable reflecting whether annotation was not given by the sonographer (unlabelled), or whether it allowed the image to be identified as a left ovary or right ovary.
 - Setting of a categorical variable reflecting what the reviewer thought had been measured in terms of:
 - Bowel – a section of bowel had been mistaken for an ovary
 - Muscle – some part of a muscle had been mistaken for an ovary
 - Vessel – the ovary was not imaged, but the iliac vessels were visible.
 - Other – some other structure had been mistaken for an ovary
- **Not convincingly an ovary:** information aggregated from the images used for measurement were assessed in terms of:
 - Setting of a categorical variable reflecting whether annotation was not given by the sonographer (unlabelled), or whether it allowed the image to be identified as a left ovary or right ovary.
 - Setting of a categorical variable reflecting why the reviewer was not convinced in terms of:
 - Object measured is too small be considered as an ovary
 - The image is too indistinct to allow positive identification

- The image lacks landmarks normally associated with an ovary such as the visible iliac vessels, etc.

Notes

1. Account needs to be taken for the case of a volunteer having only one ovary. For this reason, ovaries were called 1 and 2 rather than left and right such that a volunteer with one ovary would only have details for ovary 1.
2. One or two images may be used to measure each ovary depending on the format chosen by the sonographer (see Chapter 7). Therefore exams may have between zero and four images identified as having been used to measure the ovaries. Images used to measure ovaries 1 and 2 were identified by a binary variable; 1 = used to measure ovary, 0 = not used to measure ovary
3. It was anticipated that further reviews of this nature would be performed with the results being returned to the UKCTOCS central coordinating centre through some form of data communication mechanism.

Design

Consideration was given to creating a spreadsheet for running the review and recording its results. However, such an approach was considered unappealing due to the amount of information that reviewers were required to record and the workflow of the review as suggested by the requirement analysis. Therefore a search was made for third-party software tools that might be adapted for the purposes of the review, but nothing suitable was found. For this reason it was decided to adapt the URA osImageManager application in order to develop a custom application for the review. This decision was influenced by the possibility of the application being used for further audits by reviewers in remote locations; see requirements note 3.

The osImageManager application was constructed using Microsoft Visual Studio¹⁵⁴ as described in Chapter 4. It is a Graphical User Interface (GUI) application and is written in the C# programming language. Its design was adapted in order to support the requirements of the review tool by making changes to the user interface and business logic layers of the application. The data layer design was largely

unchanged, though at implementation an Access database was used in place of SQL Server.

User Interface Layer

The user interface of osImageManager provided much of the base functionality required by the review tool in terms of displaying a hierarchical list of trial centre, the exams they performed and the collection of images for each of these exams. This was implemented as a standard Tree Control, as shown in Figure 5-2. Upon selection of a particular image in the tree control, the image is obtained from the data layer and displayed in the ezDICOM ActiveX control as described in Chapter 4.



Figure 5-2: osImageManager adapted for Image Review. The images for each exam are contained in the nodes belonging to the volunteer’s exam date. The reviewer double-clicks the exam date node to open the exam assessment dialog box shown in Figure 5-3. Tick marks indicate that the image (or exam) has been reviewed.

Changes were made to the tree control so that it could display a tick mark against all images and exams that had been reviewed; see first exam in Figure 5-2. The dialog box responsible for leading the reviewer through the exam review workflow is shown in the right hand side of Figure 5-3. The check buttons next to the ‘Add’ buttons are used to specify that the image was used by the sonographer for the purposes of measuring the first ovary; multiple images can be marked in this way.

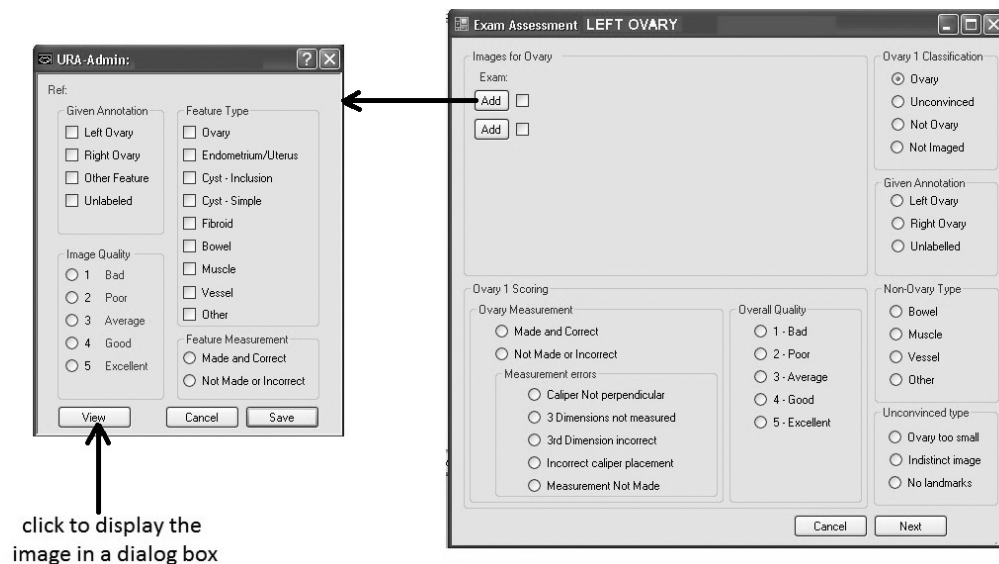


Figure 5-3: Image Review Dialog Boxes. The Exam Assessment dialog box (right) opens when the reviewer clicks on the exam date node shown in Figure 5-2. The images associated with the exam are listed in the top right of this dialog box; two are shown here. Clicking the Add button open a further dialog box (left) for recording the reviewer’s assessment of the image. After reviewing all the images the reviewer completes the Exam Assessment dialog for the first ovary before clicking the Next button to complete the Exam Assessment dialog for the second ovary.

Business Logic Layer

The business logic layer was adapted from the existing software responsible for implementing the functionality required by osImageManager when storing data in the URA as described in Chapter 4. These changes were primarily concerned with ensuring that the information in the various forms (dialogs) was complete and correct. However, it was also necessary to ensure that the objects used to store the reviewer’s information were created and stored in the data layer.

Data Layer

The data layer was also adapted from the existing software responsible for implementing the data layer osImageManager functionality of the URA. However, to support the review process two additional tables were added to the database in order to store the required data; ‘ExamReview’ and ‘ImageReview’. The ‘ExamReview’ table contained records with fields for the data stored for the left and right ovary. There was also a relational link to the ‘ImageReview’ table which contained records for the image reviews (step 1 of the review process as specified in the requirements). This was a one-to-many link so that one ExamReview record could have many

'ImageReview' records, each with information about a particular image including its association with either left or right ovary.

Implementation

The Microsoft Access DB engine¹⁵⁵ was used in place of SQL Server used in the URA. This uses the same interfaces as the larger SQL Server, but is more suited to stand-alone applications such as osImageManager adapted for image review. Therefore the application was integrated with the database simply by making appropriate changes in the database connection configuration file. The database containing the images required for the review was made by duplicating the URA database and exporting the required records. Three databases were created in this way; one for each of the reviewer groups as described in Chapter 7. The additional tables required by the adaptation of the data layer (see above) were then added to the database.

A number of laptops were setup for the reviews. The adapted osImageManager application and the Microsoft Access DB engine were installed on each of them. One of the three database files was then copied to the laptop's file system and the necessary changes were made to the database connection. In this way at least one laptop was prepared for each group of reviewers with a database containing the exams and images they needed to review. An external monitor, mouse and keyboard were provided for the purposes of the review.

The data from the separate databases was merged together after the review to form one master database. Data was then exported from this database as Excel files for analysis. Further details are given about the statistical analysis that was performed and the results obtained in the methods section of Chapter 8.

Discussion

The tool performed well during the review process which was performed in the third quarter of 2009. No significant problems were reported, though one reviewer did suggest that a small scale pilot study would have been helpful in terms of defining the requirement and fine-tuning the way data was collected. There is substance in this criticism as some of the data has not been subsequently used for analysis, so a

pilot study might have helped refine the requirements and thereby avoided unnecessary data collection.

Data from the reviews was exported from the master database in the form of spreadsheets which were initially analyzed using the open-source tool R¹⁵⁶. However, the Unit Test Framework for R described by Konig et al¹⁵⁷ was not available at the time of the research. Therefore MatLab¹⁵⁸ was used for final analysis of the data due to its better support for this type of testing. This allowed the results to be verified more robustly in terms of the scripts used to generate them.

Although at the outset it was anticipated that there would be other similar audits of TVS examinations performed by UKCTOCS, they were not conducted during the trial. Indeed the only other large-scale review of TVS exams using image stored in the URA was performed by a single reviewer in 2015 using the spreadsheet tool described at the start of this chapter.

General application for image annotation during reviews

In addition to the value that image reviews provide in terms of facilitating quality control, they also form an important component of much research involving images by creating sets of images which can be used for ground-truths^a or similar purposes. Therefore there is a need to create an application that can be adapted for use in a variety of image review scenarios, particularly those related to annotating images for the purposes of algorithm development such as that described at the end of Chapter 6. The requirements, design and implementation of such a system are described below.

Requirement Analysis

When attempting to develop software able to identify the boundaries of ovaries in images from TVS examinations, it is necessary for experts in gynaecological scanning to mark such boundaries in a significant a number of images. In this way

^a A ground-truth is a term used to describe information provided by observation. Such information may be used for training a classifier by supervised learning so has application in machine learning and related fields of study. However, the availability of a set of images with known properties as assessed by a domain expert is often a prerequisite for research; see prototyping work to automate the identification of ovary boundary discussed in Chapter 9.

the boundary found by the software can be compared to the boundary determined by the expert; see Figure 5-4. The requirements of an application that can be adapted to perform a variety of reviews involving the annotation of images can be summarised as:

- For use by multiple reviewers, though not simultaneously. The ability of the application to be used outside the central location of the trial Coordinating Centre was considered important.
- Able to operate on different datasets of images defined in terms of a list of images and their properties. The properties of these images will be imported from other sources.
- The system should have a user interface that facilitates the execution of the review. Particular emphasis was given to allowing reviewers to work quickly, track their progress and easily identify any exams they had inadvertently skipped.
- The reviewer needs to mark the image in such a way that the annotation can be separated from the original image. This suggests the need to allow the annotation layer to be separate from the underlying image layer. For example, the reviewer might define a calliper line in terms of the x and y coordinates of its start and end position. The tool needs to store this information separately from the image data so subsequent display involves first rendering the image and then overlaying a line as defined by its annotation details. In this way annotations are formed from a vector of values which can be easily compared with similar vectors generated by algorithms from features found in the image; see Figure 5-4.
- The reviewer needs to annotate images by using drawing tools to mark:
 - Calliper lines showing the distance between two points such as the largest extent of an ovary boundary in a given plane.
 - Freeform curves such as might be needed to mark the location of all points on the boundary of an ovary
 - Points at a particular location in the image. For example the correct position of a calliper which has been incorrectly positioned by the sonographer.
- The reviewer needs to add information about the image at a number of levels:

- Whole image details such as a quality score, notes, etc.
- Details about objects found in the image as defined by annotations such as the type of object (ovary, cyst, etc.) and its properties (size, classification)
- In addition to the requirement for the tool to support annotate for research, it also needs to support a range of different types of review. For example, in the case of an audit review it would be necessary to record whether the reviewer considered the callipers used to measure the ovary actually marked an ovary or not.

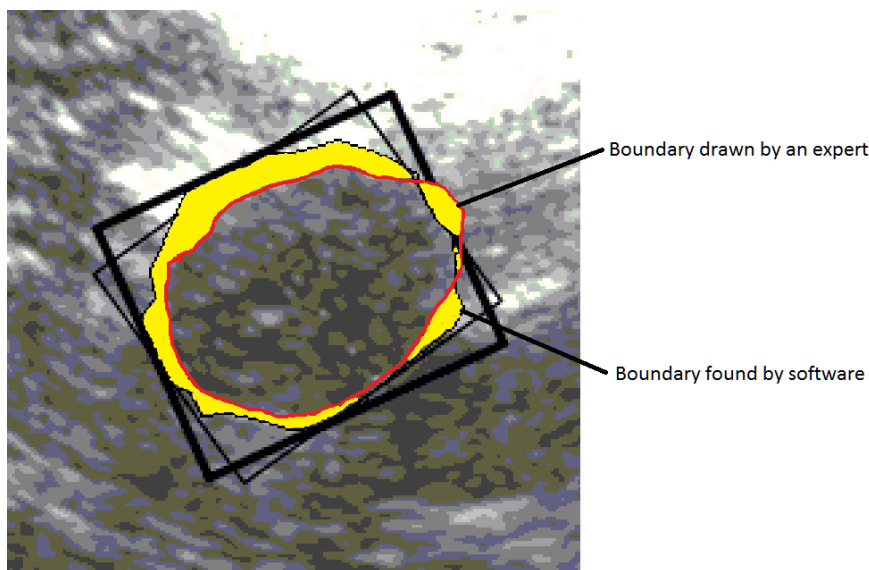


Figure 5-4: Differential Analysis of Ovary Boundary. Any error between the location of the boundary drawn by the expert and that found by software analysis of the image can be expressed in terms of a score reflecting the number of pixels in the overlapping area between the two boundaries. Essentially this is the Sorensen-Dice index, a statistic used for comparing the similarity of two samples based on the intersection of the samples. This allows the mean scores achieved by different processing runs for a given dataset of images to be compared for the purposes of algorithm selection and parameter tuning. The process of finding the overlapping area between the software and expert drawn boundaries is facilitated by defining them in terms of a sequential list of values having the same type and semantics – i.e. a vector

Design

A search was made for third-party software tools that might satisfy the requirements, or be adapted to do so. However, no suitable product was found. This reflects the novel nature of the proposed tool. For this reason it was decided to develop a new application. The adoption of a familiar set of tools and architecture reduced the

development risk and avoided the need to assimilate new technologies. The design can be described in terms of the data layer, business logic layer and user interface layer. The new application was called ImageCurator to reflect its role in allowing the curation image datasets.

Data Layer

It was decided to implement the data layer as a flat file in XML format with the images contained as bitmap files in a separate folder. This avoided the need to install a database, reducing the installation overhead if the application needed to be installed on computers outside the trial Coordinating Centre. It also simplified the development process by removing the requirement to manage the refractoring of a database due to schema changes, as discussed by Ambler and Sadalage¹⁴³.

The open-source TinyXML¹⁵⁹ library was used to implement the basic operations of creating, reading, updating and deleting XML files. The library was wrapped by a number of classes that implement the parts of data layer specific to ImageCurator. These classes interfaced with the classes of the business logic layer as described below.

An XML file called an Image Dataset File (IDF) was developed based on previous work as described in Chapter 4. It can be described in terms of its header and body. The header defines the structure of the image dataset in terms of its properties at various levels:

- Group properties that apply to collections of images in an exam; for example the identity of the exam given as the combination of volunteer reference and scan date. Other group level properties may also be defined such as ovary dimensions recorded by the sonographer which apply to more than one image.
- General image properties that apply to all images; for example the identifier of the job used to process the image – i.e. the IJF as defined in Chapter 6
- Image properties that apply to a particular image type; for example an image type was defined for TVS images containing the longitudinal and transverse sections of an ovary (LeftOvaryLSTS) with proprieties such as ‘scale’ determined by analysis of the image. It also contains a list of objects

associated with this image type such as a calliper line, a class of abnormality as specified by the IOTA group¹³, etc.

- Image object properties that apply to a particular object for a given image type; for example the dimensions of a calliper line object contained within a LeftOvaryLSTS image type.

Although the above descriptions relate to TVS images, it should be understood that support for other types of images could be implemented by changing the definitions in the IDF. The body of the XML file contains a collection of elements corresponding to image groups, each of which contains a set of group properties as specified in the header. Nested within each group element is a collection of image elements each of which has a set of general image properties as well as a set of properties specific to the image type. Furthermore, nested within each image element is a collection of image object elements, again each with its specific properties as specified in the IDF header. An example of a section of the IDF body is shown in Figure 5-5.

```

<Body>
<ImageGroups>
<ImageGroup Name="VolunteerRef=15039845,ScanDate=2008-01-07" VolunteerRef="15039845" ScanDate="2008-01-07" Notes="..." Done="n">
<Image Name="1.2.826.999999.1.9999.1.20080107011904673.1.2.0" ImageTypeName="LeftOvaryLSTS" D1mm="0.0" D2mm="0.0" D3mm="0.0" Scale="0"
<ImageObjects>
<ImageObject Name="AD1752CA-E2F4-4756-BC86-28D898711F58" Type="IOLine" ImageObjectDef="L0LSTS LS D1" Notes="..." Ax="535" Ay="160" Bx="486" By="218" />
<ImageObject Name="9E89C876-8B5F-4E38-9701-B91234E0B839" Type="IOLine" ImageObjectDef="L0LSTS LS D2" Notes="..." Ax="490" Ay="173" Bx="538" By="201" />
<ImageObject Name="B08A0126-9FB9-4888-8B66-652DC972F1E2" Type="IOLine" ImageObjectDef="L0LSTS LS Scale" Notes="..." Ax="2" Ay="114" Bx="2" By="190" />
<ImageObject Name="84E3B5ED-7118-4D10-B54E-7C801F457383" Type="IOLine" ImageObjectDef="L0LSTS TS D3" Notes="..." Ax="243" Ay="138" Bx="184" By="225" />
</ImageObjects>
</Image>
</ImageGroup>
    
```

Figure 5-5: Part of an Image Dataset File (IDF). The IDF is an XML file so properties of the individual images in the dataset are given as attributes of the corresponding ‘Image’ element as well as being contained in nested ‘ImageObject’ elements. In this case the image has been annotated with objects showing where the reviewer thought the calliper marks should have been placed, represented here by ImageObject elements.

Some values can be stored in the IDF file using standard XML data types; strings like those used for image identifiers, or integers like those used for calliper position. However, a special data type was defined for storing vectors of locations such as those used to draw the points that make up an ovary’s boundary.

Business Logic Layer

The business logic layer of the application is concerned with marshalling the data layer objects as required by the current state of the user interface. For example, if the user selects a particular image by clicking on an item in the user interface’s tree

control, then the business logic will instantiate the necessary data layer objects necessary to display it and then use their methods to render the specified image in the display, set its properties and so forth.

User Interface Layer

The user interface layer is constructed from standard controls provided by the Microsoft Framework Class library as shown in Figure 5-6. The main elements of the user interface are:

- **Tree control** – the examinations in the dataset and the images they contain are displayed in a tree hierarchy as shown on the left-side of Figure 5-6. The currently selected image is indicated by a red dot.
- **Display window** – the bitmap corresponding to the selection in the tree control is rendered in the central area of the application's main window by a standard image control.
- **Properties window** – the properties of the selected image are contained in the appropriate elements of the IDF file as described in the data layer section (above). These are displayed in the standard property control as shown on the right-side of Figure 5-6. The form of the property window is specified by the IDF file, so that different properties are displayed according to the image selection. In this way properties for images that are not from TVS exams may also be displayed.

The user may annotate the selected image by selecting a drawing tool from the menu or toolbar. Different types of drawing tools are available according the type of image selected and its definition in the IDF. To start the annotation the user presses the left mouse button and moves the mouse pointer to trace around the ovary boundary (if the computer has a touch screen, then a pen might be used for this purpose instead). Upon completion of the drawing the user releases the mouse button and this creates a new image object containing the vector of locations used to draw the points that make up an ovary's boundary. These are then stored in the IDF by serialising the object using appropriate data layer objects. The new image object is selected in the image (shown as a bold line) and its properties are given in the property window.

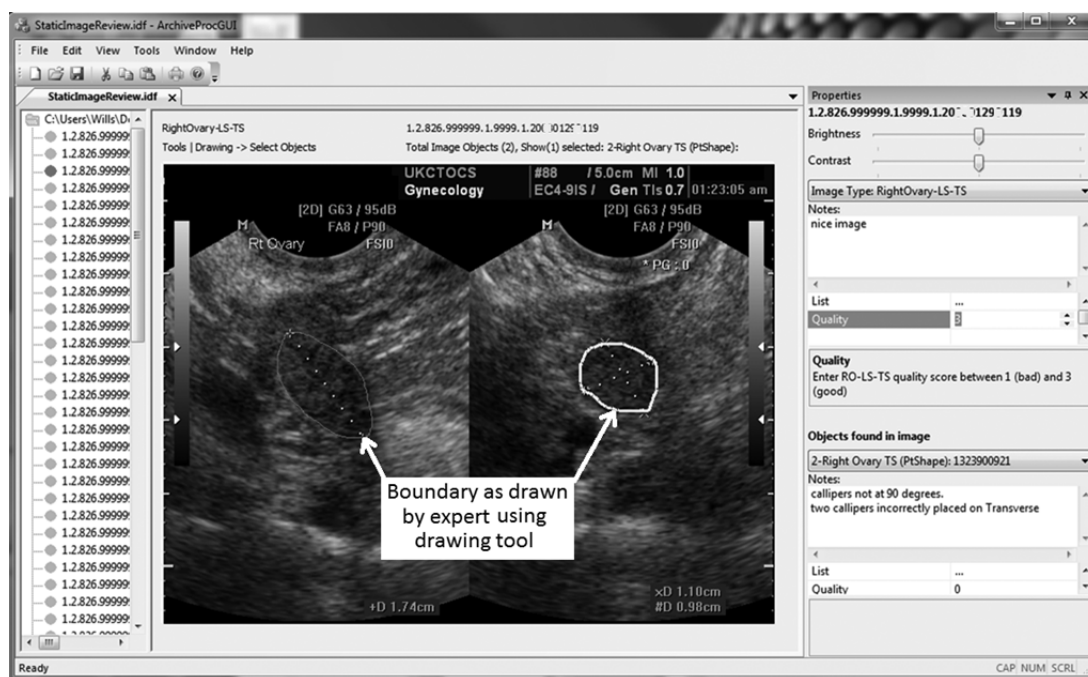


Figure 5-6: ImageCurator User Interface. The expert reviewer selects the image in the tree control (left) and then chooses a drawing tool from the menu. This allows the boundary of the ovary to be drawn on the image by tracing around it using the mouse. The reviewer can also enter annotations for the image and the boundary object by entering information into the Properties window (right).

In addition to allowing the user to add properties to selected images as outlined above, the user may also import properties for images (or exams) in the dataset using a CSV file created by the TMS or other system. In this way the information associated with a dataset can be enriched. The name of the tool, ImageCurator, reflects its role of increasing the value datasets in this way. It is anticipated that a biobank of the type proposed in Chapter 4 might be used to store collections of IDF files for the UKCTOCS image archive in order to add further utility for researchers as well as other types of users.

Implementation

The first version of ImageCurator was completed by 22 July 2012. It was installed on a PC in the UKCTOCS coordinating centre together with a dataset of bitmap files and associated IDF file. An instruction manual was also prepared. The system was used by an expert in gynaecological scanning in a small-scale pilot study on 23rd November of the same year. This was intended as a preliminary for a study of bulky ovaries. The pilot study involved reviewing images from 44 exams and took three

hours. It was judged successful, though the entry of data into the property windows was not as convenient as was the case for the type of custom dialog in the 2009 audit tool, as described earlier in the chapter. Although it was decided to proceed with the bulky ovary study, the image review was not started and the study was finally abandoned in the second quarter of 2013. The tool was not used for other research.

Discussion

ImageCurator is a novel tool which has potential for a variety of uses in the context of the UKCTOCS image archive and similar image collections. In addition to its use as a tool for annotating images such as those in the bulky ovary study, it could also be used for the type of quality control audit previously described. However, such use would require further development of the user interface to improve its usability for people who do not have extensive experience in using modern software. It may also require further development of the data layer to support use of the tool in remote locations.

A further potential application for ImageCurator is the production of image datasets for training sonographers. Such trainees might use the application for reviewing images from the URA with datasets created to contain a variety of conditions likely to be encountered during scanning. In this respect it would serve a similar purpose to the 'Atlas of Ultrasound in Obstetrics and Gynecology' by Doubilet and Benson¹⁶⁰.

Summary

This chapter has described the development of three tools to facilitate the reviewing of images from UKCTOCS. It started with a spreadsheet, which is an appropriate solution for a simple review involving one reviewer making only a few types of assessment on a collection of images. However, the need to support more complex reviews, like the audit involving assessments of many aspect of a TVS exam by a team of reviewers, required the development of a custom application; albeit one based on the existing osImageManager software. In order to avoid the need to develop new software to meet the needs of each review commissioned for study a general solution was developed that could be adapted for different types of review by means of a configuration (XML) file. Although this general solution, ImageCurator, was successfully used in a pilot study it was not used for large-scale review or audit. However, there is potential for its further development which is planned for implementation in 2017.

The main application of a tool like ImageCurator is the creation and annotation of datasets for image research purposes. In this respect the tool's ability to allow clinicians to curate a set of images representing the range of objects that need to be detected by software gives some advantage, especially as the annotation are in a form that can be easily compared with software generated values; see Figure 5-4 and related discussion. The way the IDF files generated by ImageCurator could be used in a workflow to automate the image processing associated with such a task is discussed in Chapter 6.

Chapter 6: Image Processing Workflow Tool

Introduction

The collection of Quality Control (QC) metrics for TVS exams by routinely analyzing the images generated for ovary measurement is deemed impractical when done by manual review involving experts in gynaecological scanning, except in the case of small-scale reviews such as those described in Chapters 8 and 9. However, if the process could be automated then it may be possible to gather QC metrics regularly for the millions of TVS exams that might be performed annually in any large-scale screening programme for ovarian cancer. This thesis investigates the viability of such an automated approach to obtaining QC metrics for TVS exams. In this respect the development of the type of image processing workflow tool described in this chapter would be a key part of any solution. Furthermore consideration is given to the more general problem of research involving the analysis of images and how the tools developed by this thesis project might encourage people with specialist skills to form cooperative teams.

The chapter starts by identifying the requirements of an image processing workflow tool and then reviews existing work that might facilitate its design. This provides the necessary context for the design of the tool which is then described in terms of a) its components containing processing functions, 2) software to provide a configurable workflow and c) the user interfaces necessary to make it a productive tool. The chapter concludes by considering the use of a tool in the research performed by the thesis project as well as its potential for more general research involving large image datasets.

Requirements Analysis

The requirement analysis started by considering the scope of the problem and this suggested the type of software tool that needed to be developed. The requirements of such a tool were then specified as well as the way it might be integrated with a novel approach to research involving images. This allowed the design of a tool that not only satisfied the immediate needs of the thesis project in terms of collecting quality control (QC) metrics from the images of TVS exams performed by UKCTOCS, but also allowed the possibility of further extension for the purposes of other research involving large-scale image analysis.

Problem Scope

The Quality Control (QC) of the TVS exams performed by UKCTOCS was based on measuring the visualization rate (VR) of individual sonographers and trial centres, as described in Chapter 3. For the purposes of this thesis, VR is the percentage of exams for which the dimensions of both ovaries are recorded in the Trial Management System. This indicates that both ovaries were seen during the examination as such measurements can only be obtained by the sonographer creating a static image and placing calliper marks on the object thought to be an ovary, as described in Chapter 1. Because these images are typically archived in the URA, there is an opportunity to obtain an independent and objective measure of VR by using a software tool to analyze these images as proposed in the thesis Problem Statement.

UKCTOCS performed 300,027 annual TVS only scans during the trial and 216,152 (72%) of these exams have their associated static images stored in the Ultrasound Record Archive (URA), as described in Chapter 4. However, no record was made of the purpose of each image except for non-standardized annotations made within the images themselves such as 'RO'^a to indicate right ovary. Consequently in order to obtain QC metrics for any exam it would be necessary to identify the images used to measure the ovary.

^a The image annotations were not standardised by UKCTOCS so they appear at various positions within the image and have a variety of forms such as RO, R, Right, Rt, etc.

When performing a study of images associated with a small number of exams it is possible to sort the images of interest from the rest by manually inspecting the annotations within each image. For example, the images used to measure the left and right ovaries typically consist of two images, each containing views of the longitudinal and transverse section of an ovary which may be identified by someone finding the annotations used to define left and right ovary within the image. Clearly, this type of manual sorting is impractical when studies are performed which involve thousands of exams. For this reason a computer system capable of processing large numbers of images using some form of workflow was considered necessary. The requirements for such an image workflow system are considered below.

Requirements for an Image Workflow

The requirements of the image workflow computer system are expressed in terms of the features needed to define image collections, the type of processing that may be performed on them, the support needed for high volume processing and way in which the data might be gathered for research. In addition the common problems faced by people performing research involving images are considered.

Definition of Image Collections

The workflow needs to process various collections of images extracted from the URA according to specified criteria; for example, all images from annual TVS exams classified as normal performed between 1st January and 31st December 2008. Such an image collection forms a dataset which the workflow needs to process as a unit in order to generate a set of results. There is also a requirement for the workflow to integrate with the tools described in Chapter 5 by allowing the definition of such a dataset not just in terms of the list of images it contains, but also the properties of these images at a number of levels:

- Properties unique to an individual image, such as its DICOM reference or a list of objects contained in the images (each with their own properties).
- Properties shared amongst a group of images, such as an examination reference common to all images gathered in respect of a particular TVS examination.

- Properties common to all images in the dataset, such as annual TVS exams performed within a given date range.

Properties from the image dataset will provide inputs for the workflow processing and support is needed for the basic types of binary, numeric and string data with the possibility of defining a range of acceptable values to permit validation prior to processing. An image dataset requires flexibility in terms of the definition of its various properties and list of images so that the workflow can support the needs of the different studies that will be designed during the course of a research project. There is also a requirement for the dataset to accumulate properties as a consequence of processing by the workflow so that the results of one execution run can be used to provide properties for the execution of the next.

Processing

The key function of the workflow system is iterating through the list of images in a dataset and causing the execution of a sequence of processing steps on each of them. There is a need to build such processing sequences in a flexible way to accommodate the requirements of the different studies that will be planned as the research evolves. It is also necessary to allow flexibility in the definition of each of these processing steps in terms of its inputs, outputs and the operations performed on the data (image). This suggests the need for the workflow to provide integration with software components supporting a common interface. In this way it would be possible to perform some form of novel processing on an image simply by developing a new component able to perform the required operations without the need to extend and rebuild the entire workflow system. This approach would also allow the reuse of existing functionality when defining the sequence processing steps for different studies encouraging the development of general components as well as those with special purpose functions.

Support for Imager Processing Transactions

The operation of the workflow system shares some of the requirements of a Transaction Processing Monitor (TPM) initially developed for IBM System/360 mainframes¹⁶¹ as described by Sessions¹⁶². If the sequence of processing steps applied to each image in a dataset is considered to be analogous to a transaction, then

the workflow serves to process transactions. This is a particularly useful abstraction in the case of datasets containing thousands images which may well take many hours to process and require the efficient sharing of resources like memory, database connections, etc. It also allows the image workflow system requirements to be expressed in terms of a multi-tiered architecture such that an application program in the client tier controls a program in the middle tier that executes transactions on images (and their properties) as contained in the data tier. Stating the requirements in this way encourages the development of an architecture that is flexible, scalable and robust. However, account must be taken of the complexity associated with a full-featured TPM which is mostly not necessary for the type of work being undertaken by this thesis project^b.

Data Storage and Access

The images are stored as Binary Large Object (BLOB) datatypes in records of a relational database (SQL Server) as described in Chapter 4. However, for the purposes of the studies performed as part of this thesis (see Chapters 7, 8 and 9) the images of a given dataset were exported to a local file system as bitmap files and a XML file was created to list them as well as their property values. Therefore the data layer of the workflow system needs to consider support for data storage and access in terms of both a relational database and local file storage. It is possible that other types of data store will be used in the future. In terms of storing the results of processing either a standalone file or a database could be used.

Image Processing Engine

Imagining the processing of an image dataset as a series of transactions each involving a particular image allows the part responsible for executing the transaction to be considered as an image processing engine. Although a specific need to support multiple processing engines was not identified, the possibility of new engines being instantiated at run-time to share the processing of the image dataset amongst different CPU cores (or machines) does encompass the general requirement for a

^b The origins of TPMs can be traced back to a software product in an airline reservation system that allowed thousands of terminals to access a backend database without the need for each one to have its own database connection – i.e. it introduced a middle-tier to client-server architecture to solve the problem of sharing a limited resource amongst many client programs. Although an image processing workflow may only have one client program (see Workflow Client) it does need to process thousands of transactions each processing a different image in a common way using shared components.

system that could be scaled to perform complex processing on large image volumes with high through-put.

At start-up the image processing engine needs to acquire the resources needed to execute a series of image transactions. This would mean acquiring a database connection should image access or the storage of results involve reading or writing records to the relational database. It also involves acquiring other system resources such as memory, components needed for implementing the process steps and connections for transaction queues, data logging, etc. It is the responsibility of the image processing engine to share these resources amongst the transactions it creates and release all such resources upon their termination.

The image processing engine needs to iterate through the list of images in the dataset. Conceptually this can be considered in terms of reading items from a message queue such as might be implemented by IBM's MQ¹⁶³ or Microsoft's MSMQ¹⁶⁴, but it could also be implemented in terms of a XML file (or similar) created by a transaction control programme for a given instance of the image processing engine. Each item in the message queue needs to contain the information needed to complete the image transaction, for example the reference of the image in the dataset that is to be processed. The image processing engine also needs to implement a mechanism for reading this information from the head of the queue at the start of each transaction. In the case of a message queue like MSMQ this is simply a matter of removing the first item from the queue, but other types of queues might need other mechanisms.

The main responsibility of the image processing engine is the correct execution of a transaction, usually stated in terms of the acronym ACID; atomic, consistent, isolated and durable. This means that the workflow either processes an image completely or fail completely; it is atomic. A transaction must also generate a result that is valid according to rules such as the permitted range of values; it is consistent. Furthermore transactions must operate independently of each other so partly processed images cannot be shared by different transactions; they are isolated from each other. Finally, the results of a completed transaction must be saved and not subject to subsequent alteration or deletion even if the image processing engine crashes; it is durable. However, the strict requirements of ACID may need to be

reconsidered if the workflow system was used for large scale simultaneous processing of multiple datasets in a database system. For example transactions might be allowed to complete even though the results were inconsistent so long as there was a mechanism for ensuring eventual consistency.

An image transaction can be considered as a sequence of processing steps each having given inputs and outputs for functions implemented by components. This might be realized by using a XML file to define a collection of steps in a given sequence, each with input and output parameters specified as well as the name of both the function and the component in which it is implemented. The image processing engine could use this type of Processing Steps XML (PSX) file as follows:

1. Initiate the transaction by performing any transaction start-up tasks such as identifying the next image from the dataset queue
2. Obtain the input parameters needed for the first step of the process by obtaining values from image's properties in the dataset with names that match those specified for the corresponding step in the PSX.
3. Invoke the function (loading the named component if required) so executing the code in the component using the input parameter values obtained from the dataset in step 2.
4. Copy the output parameters returned from the call and use them either as inputs to the next step or as outputs from the transaction for recording in the results.
5. Repeat steps 2, 3 and 4 for each of the processing steps specified in the PSX.
6. After the completion of the last processing step record the results before performing any transaction shut-down tasks and marking the transaction as complete.

The engine would continue processing image transactions in this way until the dataset queue was empty – i.e. all images processed. Although, the image processing steps could be defined using mechanisms other than XML there is a need for them to be formed and altered easily so the workflow system can be setup for a variety of studies involving different types of image processing.

An important part of transaction processing is the proper handling of errors so that problems in any one step of the processing result in complete failure of the associated transaction (atomic), but do not cause termination of the image workflow engine. In this way the entire dataset can be processed by repeatedly running the workflow for all transaction jobs not completed until there are no more errors. Clearly there is a need to fix (or remove) particular images that are causing errors at the end of each run, but this is much more efficient than the non-transactional approach of just processing all the images time and time again until the run can be performed without any error – particularly in the case of runs that take many hours to execute.

A further feature required of the image processing engine is the ability to trace the execution of the workflow for the purposes of investigating problems. This is closely related to the requirement for a logging feature to record the processing performance. Both tracing and logging are implemented in similar ways; they differ primarily in the way the information is used. Tracing is generally used by software developers to help find parts of code that do not operate correctly. For example, a function that returns an error code when presented with an image that is recorded in an invalid format. Logging is usually used by people who maintain the system (SysAdmins) to help them identify issues like poor performance, low memory conditions and so forth.

Tracing and logging require the software to be instrumented in some way. In the case of tracing, special lines of code are typically added to the software by hand at key points in order to record not only their execution but also the value of key parameters. These special statements are only activated when the system is built for the purposes of debugging; they are omitted for builds intended for production use. In the case of logging the same type of special statements can be added by hand, or alternatively special blocks of code (probes) can be automatically injected into the software binaries at particular locations such as just after the entry point of a function or just before its return point. Unlike tracing statements, logging statements may be retained for both debug and production builds and usually they have a facility to change the amount of data being stored in the log according to a parameter that can

be configured at runtime. This would allow SysAdmins to collect logs with various degrees of detail about the way the workflow is performing.

An image workflow system needs to allow both tracing and logging. Tracing is needed to help debug components and is particularly important as some errors will only occur when certain types of image are processed which makes it tedious to step through each transaction with a debugger for until the one with the problem image is encountered. Logging is needed to monitor the performance and health of the system. It helps identify problems such as low memory situations arising during the processing of some image datasets. This might signify the need to install more memory into the computer or change the software to perform the processing more efficiently.

In addition to tracing and logging features the image processing engine should allow the saving of images after the completion of each processing step to help with the debugging of particular image datasets. This allows a researcher to observe the impact of individual steps on the processing of the image so adjustments can be made to the parameters of functions implemented by components. However, the creation of these images needs to be under the control of the operator (see next section) so higher performance can be obtained when they are not required.

Typically TPMs are concerned with implementing security so certain types of transactions can be accessed only from client software operated by people who are logged-on with the required security tokens. However, the image workflow system does not have a requirement in this regard as the client layer software is concerned with the control of image transactions running in a workflow. Therefore it is assumed that anyone with access rights to use such client software is permitted to create and run workflows without restriction on type of processing performed by the related components.

Workflow Client

The requirements for the software that controls the image processing engine can be described in terms of the client layer program, though it serves a different function than a typical client of a TPM which acts like a terminal and allows the user to initiate individual transactions such as a booking a seat on airplane (see footnote b).

The main functions of the workflow client are coordinating the processing performed by the workflow engine and providing a user interface for the researcher.

In the simple case a researcher might use the client to specify an existing image dataset and an existing set of processing instructions (PSX) before starting the workflow engine to cause the execution of a series of image transactions, each of which might process the source image to obtain some properties or transform it in order to create a new image. Such a requirement could be satisfied by a crude command-line user interface (CUI) that allows the user to pass some parameters to the program that implements the workflow engine which is then run and stores its results in the local file system. Progress and errors messages could be output during processing. Additional commands could also be implemented so the researcher could terminate the workflow in case of error, or restart it to execute jobs not completed as a consequence of a transaction failure in a previous run.

A more complex requirement for the client might involve splitting the image dataset and then passing the various parts to workflow engines executing on different machines before assembling the different results together; an approach similar to MapReduce¹⁶⁵ as used by systems like Apache Hadoop¹⁶⁶. Further complexity could be added by installing a collection of workflow clients on the laptops of a research team so they could each process different image datasets with different PSXs using a group of workflow engines installed in a computing cloud like Amazon Web Services¹⁶⁷ or Microsoft Azure¹⁶⁸. The additional complexity of such requirements may dictate the creation of a graphical user interface (GUI) in order to shield the researcher from having to master all the commands needed to orchestrate this sort of distributed processing environment. In this way researcher might just open a webpage, select the required image dataset as well as the PSX file and then initiate the processing which could be observed progressing in some form of message window.

In addition to collections of transactions running jobs in a batch, it would also be useful to run an image transaction singularly as this would help debug the PSX. It would also help researchers understand the impact their progressing instructions had on the images, particularly in the case of images that were different in some way from others in the image dataset. For example, the first instruction in the PSX might

analyze the source image to obtain a measure of contrast to use as a parameter to a second instruction that transformed the image by increased or reduced the contrast so that the third instruction might operate on images with uniform contrast. In such a scenario it would be useful to try the instructions on a few images representing the range of contrast found in the image dataset. It would also be useful to compare the images produced after each processing step, even if these intermediate images did not need to be stored when the transactions were run as a batch. This implies that the component implementing the corresponding functions would be made aware that the transaction was being run singularly (debug mode) so would know to save the image before completing the call. However, when the transaction was run as a batch (production mode) such saving of intermediate images would not be required so giving higher performance.

The workflow client's user interface could be designed to facilitate not only the type of debugging described above, but also the creation and editing of both the image datasets and PSX. In terms of creating and editing the PSX it is necessary to represent the instructions in a manner that is understandable to a researcher who does not have computer programming skills, as described in more detail under the heading 'Design of the Workflow's Graphical User Interface'.

Gathering Research Data

The image workflow system needs to perform a given sequence of processing on all the images in a dataset and record the results. This is akin to the running of an experiment in typical science methodology. Therefore the image workflow system needs to support the concept of experiments in terms of the way its results are recorded. In addition the system should allow a researcher to create a project as a unit of organisation for related experiments, of which some may be created by other researchers in his/her team. The workflow client needs to present its users (researchers) with the list of projects for which they have access and within each project a list of related experiments so that it is easy to review the results as well as the images in the associated datasets and the instructions (PSX) that were used to process them. Consideration should also be given to the reporting requirements of projects in order to facilitate the writing-up of papers and reports. In this regard it is important to identify the exact versions of the software used in any given experiment

(including the components) as well as the image dataset and PSX used. This will allow the experiment to be repeated to ensure the consistency of its results.

The image workflow system should support the archiving of projects, experiments, image datasets (definitions as well as image files) and processing instructions (PSX) as well as the components and the system software. In this way it should be possible to recreate the exact environment and configuration required to obtain consistent results from the repeat of any experiment even if performed with different hardware and at a different location.

Common Problems of Research Involving Images

The investigation of requirements for an image workflow system revealed that researchers involved in the processing of images require a variety of skills and experience. In addition to their skills as a scientific researcher they need advanced technical abilities in:

- Building and running complex software tools such as the image workflow defined above. These tools may be developed specifically for their work, purchased from commercial vendors, or products of open-source projects (new or existing).
- Development and use of image processing algorithms. Although many libraries of algorithms are available commercially or as open-source projects it can be difficult to build them in a form that allows tool integration. In addition it may be necessary to develop custom algorithms de novo or implemented from published work.
- Imaging domain associated with their area of research. In the case of this thesis it was necessary to gain an understanding of ultrasound imaging (Chapter 2) as well as the way images were created during TVS scanning (Chapter 1).

It seems that progress in the field of image processing is being limited by the availability of people with the wide range of expertise required to perform the work. One way of addressing this problem would be to encourage people to specialise in particular areas of expertise and build an environment that allows them to work together efficiently.

The environment proposed in Figure 6-1 envisages an image research team composed of different types of specialists each with a particular function; other approaches are discussed in Appendix J. The Engineers would be responsible for building and maintaining the type of image workflow previously defined. This would provide standard interfaces so that Computer Scientists could develop their algorithms in components. These components could then be used by Image Researchers to perform the necessary processing by specifying them in some sort of script or domain specific language, as detailed later. Furthermore the tools described in Chapter 5 might allow Clinicians to perform the necessary review and annotation of images in datasets in a way that could be easily created and shared amongst the team. This would allow algorithms to be tested using a range of images considered representative of those likely to be encountered in clinical practice. Image Researchers would also be able to compare the results of image transaction runs with reference images annotated by clinical experts, so speeding-up the development of the instructions needed for their successful processing.

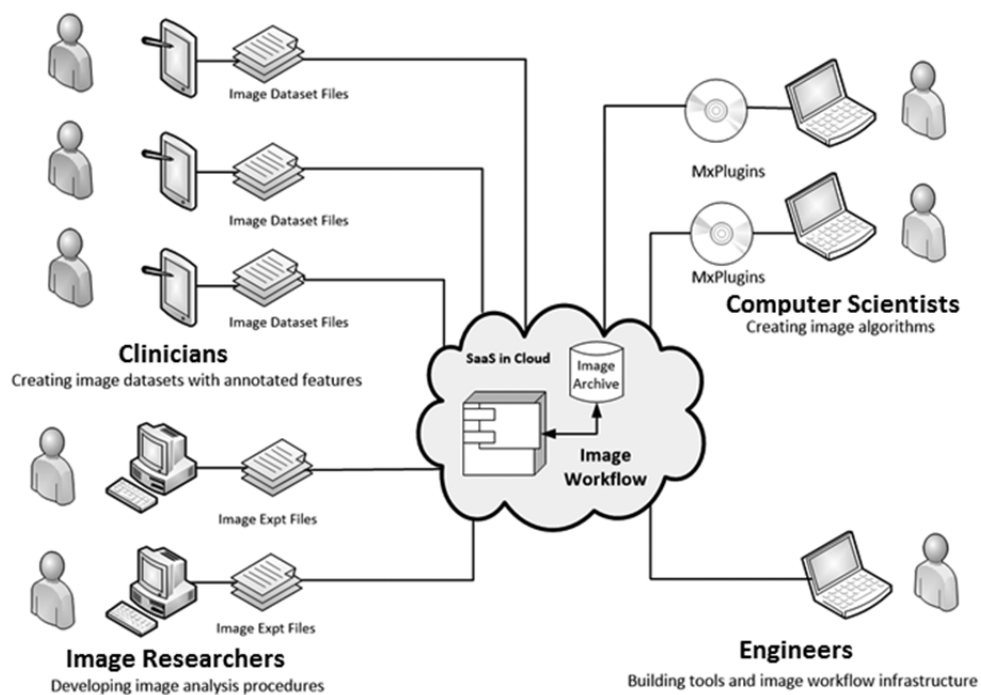


Figure 6-1: Environment for specialisation in research involving image analysis. Researchers use the tools developed by the Engineers to specify the image processing they need with needing to write any computer code. They are also supported by Computer Scientists who develop algorithms specific to their research and Clinicians who provide datasets of images to provide examples of the subject under investigation.

Figure 6-1 summarises the type of collaborative approach to research involving image analysis. Once implemented it would mean that Image Researchers would not need to master computer programming nor would they need to develop the IT skills needed to build and operate complex computer systems, for such work would be done respectively by specialist computer scientists and engineers. Instead they could concentrate on developing the domain skills they needed to understand the nature of the images they were processing; a task that would be greatly facilitated by the ready availability of annotated image datasets. Clinicians would also benefit from this new approach in terms of being able to communicate their requirements more easily by annotating the results of experiments to indicate the types of images that were particularly problematic. In this way the development process would be faster and more efficient so allowing better solutions to clinical problems to become available sooner and at less cost.

Creating a Marketplace for Reusable Components The wide-spread adoption of a common interface for image processing components has advantages in respect of encouraging Computer Scientists to provide implementations of the algorithms described in their publications so the results can be replicated by others and benchmarked against existing solutions using standard image datasets. It may also help establish a market for such components, either open-source or commercial, so assisting Image Researchers in terms of providing a greater choice of components for their work as well as allowing Computer Scientists to claim more impact through the use of their algorithms in a much wider range of applications. In addition this would help address the current problem of people publishing papers about algorithms, but failing to provide sufficient information to allow their results to be independently verified; an issue related to the more general challenges in irreproducible research as publicised by the journal Nature¹⁶⁹.

Review of Existing Work

Existing work in the area of workflow automation, componentisation of software and ultrasound image process was reviewed so its potential for application in the design of the image workflow system might be assessed.

Workflow Automation

The challenge of workflow data processing has been addressed by products like Microsoft Windows Workflow Foundation (WWF)¹⁷⁰ or Apache Taverna¹⁷¹ and VisTrails¹⁷². Specifically, as identified in the requirement analysis, the workflow must support long running batch processes, generate reproducible results and be able to handle errors such that the failure of one job would not cause termination of the entire batch. In addition it would need to support commonly available image processing algorithms and have features to facilitate rapid prototyping of the processing steps as well as debugging.

WWF was not considered a good choice as it is designed for building business rather than scientific workflow applications. Apache Taverna and VisTrails were also discounted although they are routinely used in scientific applications because they seemed more appropriate for large scale projects. Therefore an alternative image workflow was sought that was light-weight and could support the required processing of UKCTOCS ultrasound images.

A search was conducted for academic research in the area of workflow image processing systems. Commercial and open-source websites were also searched in an attempt to identify an existing product that might be used or adapted for the purposes of the work being undertaken on UKCTOCS images. The image workflow systems identified were broadly divided between those concerned with processing a particular type of image for a given purpose and those that had more general application. Cui et al¹⁷³ characterise this differentiation respectively in terms of fixed and flexible workflow tools. Fixed workflow tools are usually easy to setup and use, but have a very specific application. Conversely flexible workflows typically require considerable effort and expertise to setup, but can be applied to many types of image processing tasks.

No fixed workflow tools were found for applications related to the processing of TVS images but, a number of flexible workflow tools were identified with potential for adaptation. The principal open-source tools found were targeted at neuroimaging applications, specifically Nipype¹⁷⁴, Laboratory of Neuro Imaging (LONI)¹⁷⁵ and Java Image Science Toolkit (JIST)¹⁷⁶. However, further investigation suggested that

their adaption for ultrasound image processing would be challenging. In the case of Nipype it was reported that the tool had a steep learning curve¹⁷⁷ and its user community would be able to offer only limited help for the development of custom modules outside the field of neuroimaging¹⁷⁸. Similar problems were identified for LONI and JIST. The main commercial tools found during the search were the Oracle Grid Engine (OGE)¹⁷⁹ and MATLAB Parallel Computing Toolbox¹⁸⁰. These products are both considered as flexible workflow tools. The OGE is essentially a batch job queuing system and is typically used on a high-performance cluster (HPC). It has no specific image processing facilities, so such features would need to be implemented separately. The MATLAB product has the benefit of an extensive image processing library which had been used in early prototype work on processing UKCTOCS images. However, the performance was found to be poor in comparison with image processing algorithms implemented in C++.

When implementing an image processing workflow a choice needs to be made between either building the necessary infrastructure platform afresh, or building upon a platform provided by a third-party workflow system; see Appendix J for examples. Typically the choice will be determined by the size of the project, its similarity to an existing implementation and the experience of the development team. For example, Nipype would be an obvious choice for a project involving half a dozen researchers requiring a neuroimaging system similar to one their development team had already built using the same platform. However, a different choice might be made in the case of a project involving a single researcher needing a system for processing ultrasound images working with a developer who had no previous experience of Nipype. Whichever platform choice is made, the workflow will need to address the problem of performing the specific processing required for the type of image being analyzed.

Componentisation of Software for Image Processing

The idea of software reuse has dominated the development of programming languages since the first attempts were made to formalise the instructions need to

operate a modern computer^c. Much of the promise of the Object-Oriented (OO) languages like C++¹⁸¹ that started to emerge in the 1980s was that they would lead to the creation of a market for extensible objects (user-defined types) packaged into reusable components that would simplify the development of large and complex computer programs. However, the problems of ensuring binary level compatibility between components and their host programs was a significant challenge. It is for this reason that C++ software components were often distributed as class library source code so they can be rebuilt with the rest of the program for specific machine architectures; an approach still common today. Indeed, even components based on more recent languages like Java¹⁸² and C#¹⁸³ only achieve binary compatibility by virtue of being distributed as bytecodes that execute on specified Virtual Machines.

Problems of C++ binary level compatibility

To realise the benefit of software components that reuse code developed by others in order to simplify the task of building a large and complex program, such components need to appear like black-boxes. In this way the problem of developing the host program can be expressed in terms of integrating a collection of back-boxes and extending their functionality as necessary to produce the required functionality. The host program developer should not need to understand the internal workings of each black-box. In fact knowledge of how the black-box works is likely to prove problematic as it encourages the building of dependencies on particular aspects of the component's implementation which may preclude the use of new and better versions. Therefore the ideal model for component distribution is a binary file with given entry points that can be invoked by the host program, for example a Dynamic Link Library¹⁸⁴ (DLL) with a list of defined interfaces for delivering the specified functionality. This approach also has advantages in respect of protecting the component developer's intellectual property as it makes the internal code a trade secret.

Developing a relatively simple main executable (EXE) that uses a large collection of complex DLLs is a particular attractive proposition for the image workflow as it suggests that additional functionality could be provided by adding DLLs in the field.

^c Plankalkül is considered the first programming language developed for a 20th century computing machine. It was created by Konrad Zuse between 1943-45 for the Z3 computer and included subroutines defined as a sequence of instructions packaged as a unit to complete a particular task.

For example, a new processing algorithm could be added just by installing a new DLL on the computer upon which the workflow is running. However, this presents two key problems for workflow code; a) finding the new algorithm and b) ensuring its binary compatibility. These issues are especially problematic in the case of components built with tools different to those used to build the workflow.

It is not very difficult to develop the workflow in such a way that it can find at runtime any new DLLs installed on the computer upon which it operates, but locating the new algorithm is more challenging. This involves searching the DLL for objects of a particular type and then identifying the name of the function that implements the new algorithm. The complication arises because C++ allows object functions to be overloaded or the same name to be reused in different namespaces though a system of name mangling which is usually hidden from application programmers. The way names are mangled depends upon the tool used to build the component, so the workflow has to find the new algorithm without knowing its name. However, this problem is insignificant when compared with ensuring compatibility between the code in the workflow and the code in the component.

Typically DLLs are developed in conjunction with each other and the main executable so that in a development environment like Microsoft's Visual Studio¹⁵⁴ they are built as separate projects in a common solution. This is a convenient way of dividing a large program into a number of components. Dependencies are specified between projects in order to ensure that a change in one project will result in all its dependent projects being reconstructed whenever the solution is rebuilt. In this way altering an object implemented by a DLL by adding a further private data type, for example, will cause all the code that depends on it to change the amount of memory allocated for the changed object whenever it is instantiated and then free this same memory during its deletion. The C++ language and the development environment hide this complexity from application programmers who often have little knowledge of the underlying process. They simply add a new data member to the class of an object in a given DLL project, give the rebuild command and then let the tools do the rest of the work.

Clearly a different approach is needed when developing a workflow which needs to load new components encountered after it has been built. Without details of the

object's memory requirements, how can it know how much memory to allocate and free? Making the component responsible for allocating and freeing its own memory forms part of the solution, but it raises another problem: how does the workflow tell the component to instantiate object when it is needed and delete the object afterwards? There are further problems to be overcome such those that arise when the tools used to build the workflow are different to those used to build the component. Even if the same tool products are used, it is likely that their runtime libraries will create some objects of slightly different size when moving from one version to another which can cause a program to crash as a result of bad memory allocation arising from one component incorrectly assuming another is sharing the same library version^d.

The problems of building C++ components that can be just dropped into programs to extend their functionality arise from the language's basic lack of standardisation at the binary level as described by Box¹⁸⁵. Not only do different tool vendors use different schemes for name mangling but they also implement language features in ways that are incompatible with each other. For example, an exception thrown by a component built with one vendor's compiler is unlikely to be caught by a component built with another vendor's compiler. In addition there are other problems like that of memory allocation performed between different components as outlined above. Existing work in providing a solution to this problem is considered in the following section.

COM Solution

Microsoft developed Component Object Model (COM) in 1993 to address the problem of building components in a wide range of programming languages, including C++. It forms the basis of many other Microsoft technologies include COM+ and Windows Runtime (introduced for Windows 8 and Windows Server 2012). However, over the last decade application programmers have been encouraged to build their components using the .NET framework and languages like

^d The source of a problem known as DLL Hell which was a particular issue for all types of programs in earlier versions of the Microsoft Windows operating system. However, it is encountered less often in modern systems after it become common practice to use manifest files so applications could specify the loading of a particular DLL version (side-by-side assembly). This is proposed as an alternative to the simple convention of changing the file name to reflect the tools used to build a particular DLL – i.e. MFC42.dll, MFC80.dll, etc.

C#. This is due to the perceived difficulty of COM/C++ development. Nevertheless, COM remains a viable solution for the development of C++ components and is certain to be supported by Microsoft for the foreseeable future.

COM is based on the idea of encapsulating objects via interfaces. It provides the infrastructure necessary for host programs to find the components in which such objects are implemented as well as providing mechanisms for object creation, deletion and extension. In addition COM allows errors to be passed between different components and addresses many of the other problems associated with software component development and their use in the construction of complex systems that need runtime extensibility.

Object Interfaces

The advantage of providing access to a COM object via an interface is that the implementation is separate and can be completely hidden. Furthermore the interface can be made language independent by defining it in a text file using the Interface Definition Language (IDL). Processing such a file with a tool like the MIDL compiler¹⁸⁶ allows generation of the details necessary for interface to be invoked from multiple computer languages. It also generates the files necessary to support Remote Procedure Calls (RPC) so an object can be invoked via its interface not only when the component is loaded within the same process (in-process) but also when loaded in different processes running on the same machine (local servers) or in processes running on entirely separate machines (remote servers). This gives significant flexibility in terms of component installation and use.

The MIDL compiler allows COM objects to support multiple languages by generating a Type Library file for them. This can then be distributed with the component and used to generate the bindings necessary for the interfaces of its objects to be invoked from general languages like C, C++ and Java as well as from Microsoft specific languages like Visual Basic and C#. Type Libraries are supported by development tools like Visual Studio and allow an application programmer to invoke a method of an object implemented by a COM component in much the same way as calling a member function belonging to a C++ class in a dependent DLL project. However, the code needed to obtain the interface and manage the object in terms of its creation and destruction is a bit more involved.

All objects created from a COM component must support an interface called `IUnknown`. This is the base COM interface and all other interfaces must either directly or indirectly derived from it. It has three methods:

- `HRESULT QueryInterface([in] REFIID riid, [out] void **ppv);`
- `ULONG AddRef(void);`
- `ULONG Release(void);`

The `QueryInterface` method is passed a unique identifier for the requested interface and returns with the address of a pointer variable containing the address of the requested interface as implemented by the object – i.e. a pointer to a pointer. The other two methods are concerned with managing the lifetime of the object as described below. If the COM doesn't support the requested interface then `QueryInterface` returns an error value, otherwise it returns zero. Each interface has its own unique identifier which is a form of global unique identifier (GUID^e) and is assigned during its initial definition as described above.

In the case of a COM component developed using C++ the address of the interface returned by `QueryInterface` is in fact the location of a Virtual Method Table (VMT or `vtable`), the first three entries of which are pointers to the functions `QueryInterface`, `AddRef` and `Release` as implemented by the object. The remaining entries of the `vtable` are pointers to other functions implemented by the object as described in its interface definition. VMT is the mechanism whereby objects can have virtual functions so allowing callers to invoke an object's class methods on the basis of its actual type at runtime rather than on the basis of any design time assumption about type. This is the basis of polymorphism^f and implies that calls to such virtual functions will use memory addresses assigned dynamically during run time rather than addresses fixed-up by the linker at load time.

COM exploits the fact that the addresses of virtual functions are dynamically assigned to the `vtable` for the purposes of allowing them to be called by code outside

^e GUID is a four byte data structure containing a value generated by an algorithm that guarantees its uniqueness

^f Polymorphism is a key aspect of object-oriented programming as it allows different objects to behave in their own way in response to a common message like 'print yourself'. In this way an object modelling a text file might print a series of text lines and an object modelling a drawing might print a series of graphic shapes.

the component – i.e. code that has no way of discovering addresses fixed-up by the linker when the component is loaded because it was developed with a different set of tools. Although not explicitly stated by the language standards, all C++ compilers on the Windows platform implement this virtual call mechanism in the same way so a given class will have the same vtable structure in all cases. In this way the calling of such virtual functions can be made entirely independent of the tools used to develop the components in which they and their associated objects are contained. However, for this mechanism to work a common protocol is needed for calling functions and passing parameters. In addition parameters must only contain data types with a universally agreed structure – i.e. plain simple types used by the C language. Therefore the development of a C++ COM component requires the definition of an interface in IDL that under the covers can be translated to a C++ interface containing only virtual functions with simple C datatypes as parameters. A component that conforms to these conventions can then be used by code developed with different tools and even languages allowing a great deal of flexibility in its application.

COM components developed in languages other than C++ will solve the above problems in different ways. However, irrespective of programming language there is a need to define interfaces using IDL as well as for all legal COM interfaces to be derived, directly or indirectly, from `iUnknown`. It is also necessary for the component's objects to be accessed only through such interfaces and for there to be no dependence between the tools employed to create the component and the tools employed to use it.

Object creation and destruction

COM components need to have responsibility for creating and destroying their own objects for only the component can be expected to know what is required in terms of its object's resource requirements, as previously discussed. Therefore after the required COM object has been created there needs to be a mechanism to allow its destruction for when it is no longer required. The creation and termination of a COM object is shown in the following code fragment:

```

1: CLSID clsid;
2: IID iid;
3: IDispatch* pDispatch;
4: HRESULT nRes1 = CLSIDFromProgID(OLESTR("ukctocs.calseek"), &clsid);
5:
6: HRESULT nRes2 = CoCreateInstance(clsid, NULL, CLSCTX_INPROC_SERVER,
                                IID_IukctocsCalSeek, (void **) &pDispatch);
7:
8: ...
9:
10: pDispatch->Release();

```

Code Fragment 6-1: COM Objects are created by obtaining a unique identifier for the object that needs to be created and passing it to CoCreateInstance in its clsid parameter. This causes the object to be created and returns a pointer to its requested interface (IID_IukctocsCalSeek). Host code may then use the methods provided by this interface to perform tasks like finding calliper locations in an image. However, after it has finished with the interface the Release method must be called as shown in line 10. In this way the object knows when all its interfaces are no longer required so it can safely destroy itself.

The call to `CLSIDFromProgID` is simply a way of obtaining a unique identifier (GUID) for the class from its human readable equivalent. In this case "ukctocs.calseek" could be the name given to the class that defines an object used to find callipers. COM support is built into the operating system, so `CLSIDFromProgID` looks-up the name in a type of database called the Windows Registry and returns after writing the appropriate identifier value into the `clsid` variable. This class identifier is then used by `CoCreateInstance` to find and load the component which is packaged as an executable Dynamic Link Library (DLL), again by looking-up information stored in the Windows Registry. Once loaded the object is created and the `pDispatch` pointer passed into `CoCreateInstance` is then initialised with the address of the vtable containing the pointers to its virtual functions. Therefore upon return from the `CoCreateInstance` call `pDispatch` becomes an output parameter contained a pointer to the object's `IDispatch` interface. It thereby provides access to the newly created COM object's methods which can be used as required (represented by '...' in line 8).

It should be noted that the COM object it is not necessary destroyed by the call to `Release` in line 10 because some of the code (line 8) may have passed the pointer `pDispatch` to another object that might want to continue using it. It is for this reason that any code wanting a copy of the pointer must first call its `AddRef` method. This

increments a use count which is subsequently decremented by each call to the `Release` method. Therefore the object is only actually destroyed when the count reaches zero. At this point no other code should be using the object. The failure of a programmer to correctly match the `AddRef` and `Release` methods will result either in the object being destroyed whilst it still might be used or leaving the object as an orphan which is never destroyed – both deeply unsatisfactory scenarios.

`IDispatch` inherits from `IUnknown` and is an example of how objects can be extended as described below. The other parameters to `CoCreateInstance` define how the object will be created (in-process) and provide a reference to the identifier of the interface that is being sought – i.e. ensures the initialised `pDispatch` points at a vtable for an `IDispatch` interface rather than an `IUnknown` interface. Definitions of such things as the types of the pointers to object's various interfaces and interface identifiers (`REFIID`) are given in a C++ header file created when the programmer added its `TypeLibrary` to the Visual Studio Project.

Once `pDispatch` has been initialised by `CoCreateInstance` a pointer to any of the object's other interfaces can be obtained by calling its `QueryInterface` method. The only other requirement for using such a COM object is the need for the host application's main execution thread^g to call the functions `CoInitialize` and `CoUninitialize`. These calls respectively initialise the COM environment for use by the program when it starts and performs any clean-up before it terminates.

Object Extension

One of the key benefits of Object-Oriented (OO) programming languages like C++ is that you can reuse the code for implementing one object by using it to create another object that enhances its functionality. For example an object that contains properties and methods that model an 'employee' might be extended to create another object that models a 'manager' – i.e. a manager is an employee, but with additional responsibilities. This is the concept of inheritance. Another approach to the reuse of code is the idea of creating a new object from a number of existing objects. For example you could build a 'car' object from four 'wheel' objects and an

^g Programs always operate in the context of an execution thread created by the operating system. If the program creates a further thread then `CoInitialize` must be called within its context before any attempt is made to use COM. Such a thread must also call the corresponding `CoUninitialize` before it terminates.

‘engine’ object – i.e. a car has an engine and four wheels. This is the concept of composition. Part of the skill of being a C++ programmer is deciding when forming new objects whether to use inheritance or composition. In general composition is preferred as inheritance tends to break the encapsulation of the base class as a consequence of the derived class having the ability to override its methods. This allows changes to behaviour which violates the principle of keeping a class open for extension, but closed for modification¹⁸⁷.

When extending a COM object you are required to use composition. This is because the internals of COM objects are hidden behind its interface(s) which cannot be altered. Therefore should an object need extending, then a new interface is defined which is implemented by writing code for the new methods and also invoking methods in the original object for the existing methods – i.e. the new object contains the existing object as well as some addition code. For example if an existing COM object has an interface with a method to calculate its purchase price and there is a new requirement to calculate its sale price, then a new interface must be defined with a sale price method which may be implemented by developing a new object containing the existing object and some new code for the new interface; see Figure 6-2. Such code might invoke the purchase price method using the interface provided by the existing object and then multiply the returned value by 1.5 to provide a value for the sale price method in the new interface that is responsible for implementing.

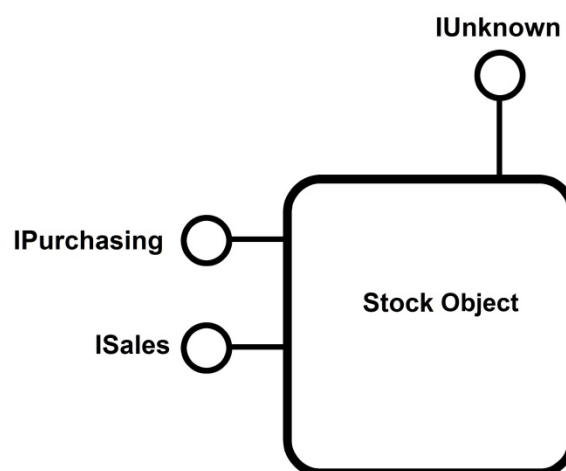


Figure 6-2: COM object is typically represented by a box with ‘lollypops’ for each of the interfaces it supports. In this case the Stock object provides two interfaces in addition to the standard IUnknown interface.

COM doesn't require each interface to be developed anew. Indeed, as previously discussed, a valid COM interface must inherit directly or indirectly from `IUnknown`. Therefore the extension of a COM object will typically involve creating a new interface that is like one of its existing interfaces, but contains additional methods – i.e. the new interface inherits from an existing one, which in turn may inherit from `IUnknown`. Commonly a COM object will implement a number of interfaces any of which can be obtained from each other using the `QueryInterface` method of `IUnknown`.

COM only allows an interface to inherit from one other interface – i.e. it doesn't support multiple-inheritance. It is also important that such interfaces are not changed once they have been defined. Consider, for example, how a program that depends on a particular method would behave upon installation of a new version of the component that had removed the method from the object's interface; clearly it wouldn't work correctly. For similar reasons the semantics of the interface should not change, so a method that returns the sale price excluding tax should not be changed to return the sale price including tax. In order to avoid such problems it is typically the case that any change a COM component involves creating a new interface.

Object Error handling

The C++ language supports the throwing and catching of exceptions as a mechanism for handling errors. This means that upon a function (method) detecting an error situation it can use the simple `throw` statement to create a type of runtime exception rather than returning an error code. The processing of the `throw` statement involves unwinding the call stack until a block of code is reached having an initial `catch` statement with a parameter for handling the particular exception type. The code in this block should then take the necessary steps to allow the program to recover from the error condition.

The intent of exception handling is to avoid the need for programmers to test for error codes returned by functions called from their code. Instead they can implement `catch` blocks to centralise error handling in the parts of their program that are responsible for error recovery. Unfortunately, programmers rarely structure their programs in such a way. Therefore `catch` blocks often do little more than `throw` a

further exception in the hope that someone else will have implemented a `catch` block that does something useful. Consequently when such programs encounter a simple issue, like not being able to connect to a database, the code spends 30 seconds or so throwing and catching exceptions before finally terminating unexpectedly, possibly with a message like “unknown error”.

The C++ language has no binary standard for the throwing or catching of exceptions. Therefore unless the same tools are used to build both the COM component and its host program, exceptions cannot reliably pass between them. It is for this reason that COM requires the simple convention of testing for an error code after calling a function in a component, or a part of the COM runtime. Therefore the return type for most methods (functions) defined by a COM interface is `HRESULT`, or similar. The implementation of such a function in the component is contained within a `catch` block so any C++ exceptions thrown by its code, or code it invokes, are caught and translated to an appropriate error code which the function then returns. In this way C++ exceptions can be used within a component, but they never transition beyond its interfaces. There is a mechanism for component code to set error information within its thread of execution, so that the host code can subsequently obtain more information about the error by invoking methods in the COM object’s `IErrorInfo` interfaces in much the same way as calling `GetLastErrorInfo`¹⁸⁸ in a traditional C program.

Creating and Using COM Components

The basic architecture of COM is not difficult to understand in terms of the way it allows components to implement objects that can be created by host code and used through interfaces, or extended by composition. However, there are a myriad of technical details to understand and experience to be gained before programmers can be considered competent at creating or using COM components, particularly in the case of people using C++. Whilst COM lies at the core of the major technologies implemented by Microsoft, the challenges of mastering it have resulted in the promotion of other approaches to componentisation such as the .NET languages and runtime. However, such options are not optimal for the development of high performance code, particularly when it needs to be integrated with a substantial body of existing C or C++ code.

OpenCV Image Processing Library

Image processing libraries are available for a variety of different computer programming languages and the interoperability (interop) of many languages means that they may be used in heterogeneous environments so allowing, for example, a Java library to be called from C++ program and vice versa. OpenCV¹⁸⁹ is an example of such a library. It is written in C/C++ and is optimised to take advantage of multi-core processing as well as taking advantage of hardware acceleration by supporting Intel's Integrated Performance Primitives (IPP)¹⁹⁰, NVIDIA's CUDA¹⁹¹ and OpenCL¹⁹². OpenCV can be hosted on a variety of operating systems including Windows, Linux and Mac OS and provides Application Programming Interfaces (API) in its native C/C++ with language bindings available for Python, Java and MATLAB. In addition, support is offered for programs written in C#, Perl and Ruby.

The OpenCV API provides access to more than 2,500 optimised algorithms for image processing and computer vision applications. Support for image processing includes functions for filtering and transforming images as well as for feature and object detection. Functions are also provided for more advanced types of image processing such as an implementation of an algorithm that uses the active contour model to find boundaries, as attempted by the prototyping work described in Chapter 9. According to its website, OpenCV has a user community of more than 47,000 people and in excess of 9 million copies of the library have been downloaded. Although now an open-source project (BSD licence), OpenCV was originally developed fifteen years ago by Intel and since then its development has been supported in varying degrees by companies such as Microsoft, Google and IBM. For these reasons it is considered to be a stable and well-developed library.

Design of a Workflow for Processing Images

The design of the image workflow system started with a review of some key decisions about how it would be built. An overview of the design was then performed to identify the main elements of the system and the way they would be used.

Key Design Decisions

Choice of Workflow System

A search of the literature and commercial websites for image processing workflows failed to reveal any systems suitable for use or adaptation by UKCTOCS. No workflows were found that had been developed specifically for processing TVS images and workflows with general application required resources that were not available to the thesis project – see Review of Existing Work, Workflow Automation. Therefore it was decided to develop a system called ‘*ImageCompute*’ that would allow the processing of large image datasets from the URA as well as having potential for use in other types of image processing, both academic and commercial.

Choice of Image Processing Algorithms

Preliminary work on the project suggested that there would be significant change in the type of processing performed on the images as the project progressed. Therefore, in addition to any novel algorithms that would need to be developed for specific tasks there was a need to use standard algorithms implemented in a software library as an alternative to implementing them anew from descriptions given in journal articles; an undertaking that would require a substantial amount of time and resources. This is because even if the source code was provided by the journal, consideration would need to be given to matters seldom addressed in publication like replication of the development environment, configuration management and testing. It is for these reasons that the use of a library containing efficient implementations of the required type of algorithms was preferred, particularly one that had been professionally developed and was well-supported. Accordingly, the main image processing library selected was the open-source product OpenCV, as previously described in the section headed ‘Review of Existing Work’.

Choice of Development Language and Environment

Whilst some languages have advantages over others in terms of factors like performance, flexibility and features there are often associated disadvantages. Therefore choice of language for a given task is usually a compromise which is biased by such considerations as the skills of the software developers and the availability of code which might be reused to reduce the scope of the work.

The availability of a good library containing a wide variety of algorithms for image processing tasks was a significant factor in language selection for the project. This was because language interop usually impacts performance which is an important consideration when processing thousands of images. Therefore C/C++ was selected as the development language and Microsoft Visual Studio was chosen as the development environment as this took due account of the native language API of its main image processing library (OpenCV) as well as the software development resources available to the project.

Choice of Architecture for Components

COM provides a reliable and effective architecture for building and using software components. However, it was not considered suitable for building the components necessary to perform the workflow image processing for the thesis project. This is because COM is a proprietary architecture for Microsoft Windows and requires developers to acquire significant skill and experience, mostly due to its need to support a wide variety of usage scenarios. These factors mean it is unlikely to be widely adopted for use in the type of image processing environment proposed by Figure 6-1 which encourages the development of algorithms in general purpose components with common interfaces for use by researchers. Therefore it was decided to develop a new architecture that provided the benefits of COM for image workflow processing in terms of tool flexibility and dynamic runtime integration, but without much of its constraints and complexity. This new component architecture is called *MxPlugin* and has now been released as an open-source product with a BSD licence hosted by Codeplex¹⁹³.

Design Overview

The main elements of the system design are formed by the plugin component architecture, the software necessary to provide and configure the workflow as well as the user interfaces required to make it useful for the purposes of research. The design of each of these elements is summarised in the following sections. However, it should be understood that the design developed in an iterative and incremental fashion in keeping with the Agile approach adopted for the work.

MxPlugin Architecture

MxPlugin was designed to provide a way of allowing *ImageCompute* to access algorithms in new components discovered at runtime. In this respect it addresses the problem of finding the names of new functions in objects whose memory and other resource requirements are only known to the components in which they are implemented; see previous discussion on the problems of C++ binary level compatibility. Therefore the MxPlugin architecture can be described in terms of:

- **Plugin components** that provide standard and extensible interfaces for host applications that can be found and invoked at runtime without any dependency on particular development tools.
- **Host applications** that use MxPlugins to extend their functionality
- **Plugin management facilities** to help host applications find and use MxPlugin components installed on a given computer
- **Support software** to simplify the use of MxPlugin components such as error handling and logging

Plugin components

MxPlugin components are implemented as dynamic link libraries (DLLs) which can be found and loaded by host applications at runtime. The need for the host application to link at build time with a static (stub) library containing information about the DLL's objects and functions is avoided by using the same sort of interfaces as previously described for COM. These interfaces are implemented in C rather than C++, so are defined as `MXSTDMETHOD` which resolves to `__stdcall`. This avoids the application of name mangling by the compiler. Such functions are also defined as pure virtual functions so that they can be accessed using the standard vtable structure. The base interface for MxPlugin is called `IMxPluginBase` defined as shown in Code Fragment 6-2.

The construction of an MxPlugin component using Visual Studio is facilitated by the provision of a Project Template which automatically creates the necessary files and provides appropriate naming according to values specified during project creation. Essentially, though an MxPlugin component is a C/C++ DLL which operates together with a statically linked library containing standard plugin management

facilities as described below. The Visual Studio template and the other files needed to create an MxPlugin component are set up on the development PC using the MxPlugin Software Developer's Kit (SDK) installation program which can be downloaded from its website – mxplugin.codeplex.com.

```
class IMxPluginBase
{
    public:
        virtual long MXSTDMETHOD DestroyPtr(void) = 0;
        virtual long MXSTDMETHOD DuplicatePtr(void) = 0;
        virtual void * MXSTDMETHOD Dynamic_cast(const GUID IID) = 0;
        virtual const GUID MXSTDMETHOD GetInterfaceGUID(void) = 0;
        virtual const char *MXSTDMETHOD GetInterfaceName(void) = 0;
};
```

Code Fragment 6-2: The MxPlugin IMxPluginBase Interface serves the same purpose as the COM IUnknown interface. It is implemented by all MxPlugin objects and provides methods to manage the object's lifecycle (DestroyPtr, DuplicatePtr) as well as to obtain pointers to other interfaces supported by the object.

Interfaces of an MxPlugin component are implemented by creating a C++ class which inherits from the interface and provides the class method code given respectively in its declaration (.h) and implementation (.cpp) files. Standard implementations are provided for the methods of `IMxPluginBase` so creating a new interface is just a matter of creating the interface's unique identifier (GUID), defining the interface in a header file as a derivative of `IMxPluginBase` (or other interface so derived), adding the new class to implement the interface and then providing the implementation of its methods in terms of either novel code or code copied from the standard implementations. MxPlugin does not use IDL to define its interfaces as it only supports C/C++ and does not allow components to be hosted by local or remote servers so has not need to support RPC calls.

An MxPlugin component, like a COM component, only permits single inheritance for its interfaces. MxPlugin objects are also be extended by composition in much the same way as COM. That is to say an MxPlugin object can use other MxPlugin objects implemented in the same or other components by constructing them as private data types and then using these objects to implement part of its own interface method(s). MxPlugin components have the same limitations as described for COM

components in terms of using standard C types as parameters, not allowing exceptions to transit interfaces and requiring matching `DuplicatePtr` and `DestroyPtr` calls whenever interface pointers are passed or copied – a requirement that may be made redundant by the use of smart pointers¹⁹⁴ in some future enhancement of MxPlugin.

The development of an MxPlugin component has an obligation to provide an installation program which has responsibility for performing all the tasks necessary to setup its files on the target computer as well as to do any required registration and so forth. The task of developing an installation program is facilitated by the MxPlugin website's provision of sample source files for the open-source installer `WiX`¹⁹⁵; software originally developed by Microsoft and still used for the distribution of some of its products.

Plugin management facilities

MxPlugin is not part of the operating system and is unlike COM in this respect. Therefore support for the management of MxPlugin components needs to be provided. This is achieved by the installation of both a software library and application; respectively `MxPluginMngr1` and `MxPluginManagerGUI`.

The software library `MxPluginMngr1` has static stub library (.lib) that should be linked to a host application in order load the associated DLL containing functions that facilitate the runtime loading of MxPlugin components and their dependencies so that their object interfaces and methods can be accessed. This library is only compatible with host applications developed with the Visual Studio 2015 tool set (hence the terminating '1' in its name) and further libraries will need to be developed for host applications developed with future releases of Visual Studio or other tool sets from other vendors. However, this is not a significant limitation as it is not anticipated that the functionality it provides will need to be changed or extended at runtime. In this respect, the library differs from the MxPlugin components it supports. At present `MxPluginMngr1` will be installed during the installation of any standard MxPlugin component, unless already present on the target PC. However, in the future `MxPluginMngr1` may be distributed with the host application and installed on the target PC (if not already present) as part of its installation program in order to encourage wider tool set usage.

The MxPluginMgr1 library provides functions to find and load MxPlugin components which can be called by the host application as described below. These functions access the operating system's environment variables to find the root location for all MxPlugin components and then iterate through the various subdirectories to create a list of all such components installed on the PC. Each of these MxPlugin components are installed in their own subdirectory which contains a text file called MxPlugin.dep specifying how to load any dependencies apart from those contained within the same subdirectory, as setup during installation. Therefore when the host application asks the MxPluginMgr1 library to load a given MxPlugin component, the appropriate subdirectory is found from the list of installed components and its MxPlugin.dep file is used to control the way the operating system will load the component and its dependencies; typically standard DLLs. The software application MxPluginManagerGUI presents users with a list of all the MxPlugin components installed on their PC as well as giving details of their vendors, interfaces, file locations and so forth; see Figure 6-3.

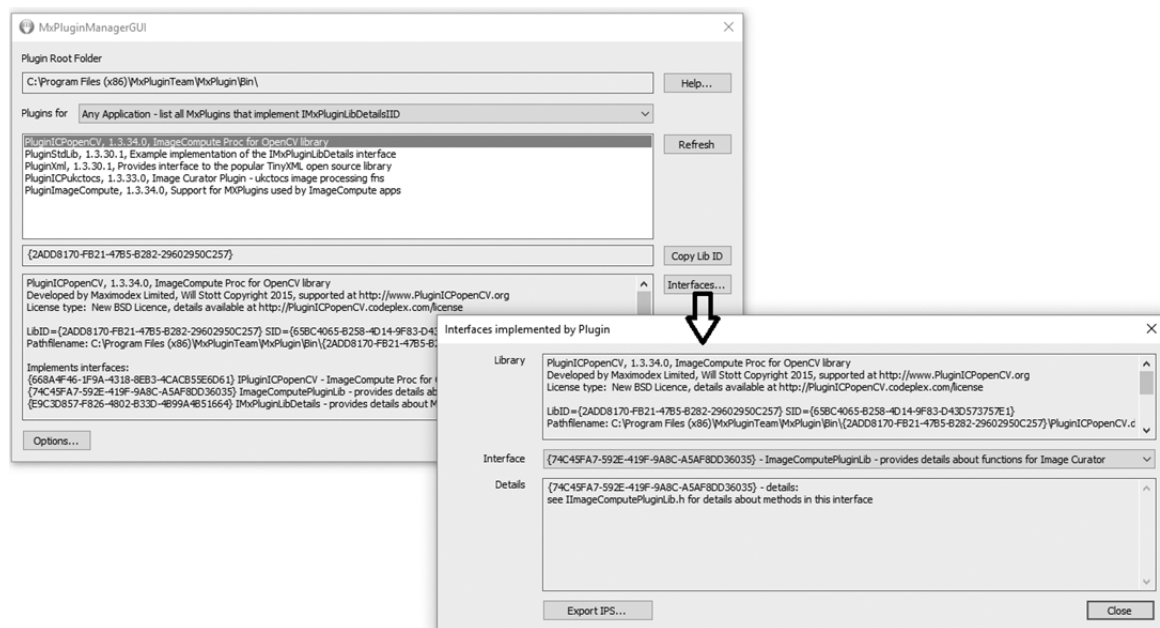


Figure 6-3: MxPluginManagerGUI allows a user to discover all the installed MxPlugin components and obtain details about what they do and the interfaces they support. The left hand dialog shows that the PluginICPopenCV component is available and provides ImageCompute functions implemented by the OpenCV library. The right hand dialog shows that this component supports the ImageComputePluginLib interface. Furthermore one of the methods in this interface allows the generation of an IPS file containing details about the image processing functions implemented by the component as indicated by the Export IPS button which appears only when ImageComputePluginLib is selected.

In addition to providing information about the various MxPlugin components installed on a PC, MxPluginManagerGUI also allows the setting of options for the tracing and logging of information generated as these components execute. In this way users can gain access to details that may help them debug the image workflow or monitor its operation.

Support Software

The MxPlugin SDK provides developers of both components and host applications with a number of utilities with the aim of facilitating their work:

- **Error Handling** – an alternative to C++ exceptions that allows the error condition to be recorded at the line of code failure together with details of variables, passed parameters and so forth.
- **Logging** – integrates with the above error handling to display error information for development or production use according to options set on the target compute using MxPluginManagerGUI.
- **Utility** – provided a number of helper functions for accessing Windows Registry, operating on files, running command line applications, conversion utilities, etc.

The above facilities are provided by MxUtils1 which is a stub library that should be linked to a host application so its associated DLL is loaded at program start-up in the same way as MxPluginMngr1.

Host applications

MxPlugin support for the development of host applications is provided by its SDK which can be downloaded from its website as previously mentioned. The key part of this support is provided by functions in the MxPluginMngr1 library (see above) which allow MxPlugin components to be loaded and their objects to be created and released as shown in the Code Fragment 6-3. It can be seen that the MxPlugin code is similar in structure to the equivalent code for loading a COM component, creating one of its objects and then releasing it; see COM Object Creation and Destruction.

```

1: MxPluginManager pluginMngr;
2: if (pluginMngr.RefreshPluginsInRootFolder (PRODID_DefaultPluginRootFolder,
                                           SID_Universal) == true)
3: {
4:     MxPluginLib *xmlLib = pluginMngr.GetPlugin (PRODID_PluginXml);
5:     if (xmlLib != nullptr)
6:     {
7:         IPluginXmlDoc *doc = nullptr;
8:         doc = static_cast<IPluginXmlDoc *>
                (xmlLib->CreateClassInstance (IPluginXmlDocIID::Guid()));
9:         if (doc != nullptr)
10:        {
11:            ...
12:
13:            doc->DestroyPtr();
14:        }
15:    }
16: }

```

Code Fragment 6-3: An MxPlugin Object is created by first obtaining a pointer to its component as shown in line 4. This pointer is then used to invoke the component’s CreateClassInstance function which creates the object and returns a pointer to the required interface – i.e. IPluginXmlDoc. Your code may then use the methods provided by this interface to perform tasks like creating XML files. However, after you have finished with the interface you must call its DestroyPtr method (line 13). In this way the object knows when all its interfaces are no longer required so it can safely destroy itself.

MxPlugin Image Processing Components

MxPlugin components provide a general solution to the problem of componentising C++ code in a way that provides binary compatibility. However, in order to provide a solution that allows extensibility for the envisaged image workflow processing, the general component interface given by `IMxPluginBase` needs extending:

- **IMxPluginLibDetails** - provides details about functions for Image Curator
- **IImageComputePluginLib** - provides details about MxPlugin and its interfaces

Therefore MxPlugin components used for image processing in the context of the proposed image workflow must implement the above interfaces so they can be accessed in a standard way by the workflow software. MxPlugin components with these interfaces are termed `MxPluginsForImages`. To reduce the amount of work necessary for `MxPluginsForImages` components to implement the required interfaces

with the workflow software, a special MxPlugin component was developed called MxPluginImageCompute. In this way MxPluginsForImages components are able to share the same MxPluginImageCompute component code by adopting the type of composition approach previously described in the context of object extension. The following key interfaces were created in this regard for use by MxPluginForImages components:

- **IImageCompute** – includes methods to allow MxPluginsForImages components to create definitions for the algorithms they implement – see IImageComputeFnDef.
- **IImageComputeFnDef** – includes methods so that the correct parameters can be passed to and from the workflow as defined in the instructions of the PSX file
- **IImageComputeFn** – includes methods so that the values for the parameters defined using IImageComputeFnDef interface the can be passed to and from the workflow.
- **IImageComputeFnCall** – includes the method necessary for the workflow to invoke the algorithm using the parameters passed to it using the IImageComputeFn interface.

These interfaces were implemented by MxPluginImageCompute so that access to the workflow software could be provided to any MxPluginsForImages component simply by it instantiating the appropriate object and then invoking the required methods. An important element in the general design of MxPluginsForImages components was the concept of making them self-documenting. In this way new components could inform their host applications of the algorithms they implement each with a description of its purpose, the parameter types they require and the acceptable range of values for them.

Creating and Using MxPlugin Components for Image Processing

MxPlugin provides an architecture for componentising C++ objects that is much simpler than COM and should be considerably easier to learn. In addition, it is feasible to implement MxPlugin on platforms other than Microsoft Windows so making it an attractive proposition for a wider range of academic researchers. It has been developed as an open-source project, so has the potential to attract a large

community of developers and users, though it currently has a low profile as to date it has not been promoted or published.

The use of MxPlugin components for image processing (MxPluginForImages) solves the problem of providing a workflow system to which new algorithms can be added by installing C++ components developed using a variety of development tools. The availability of an installation program for such components streamlines their setup (and removal) from a computer by a non-technical user. This might involve just clicking on a link provided by a Journal (or component vendor website) in order to install a new component so that its algorithms can be benchmarked against existing implementations; an approach consistent with the type of image processing environment proposed by Figure 6-1. Furthermore the interfaces provided by MxPluginForImages components allow such algorithms (and their documentation) to be accessed from a custom editor which opens-up the possibility of developing a domain specific language (DSL) for defining the processing steps performed on a collection of images.

Workflow Processing Software

The use of XML for specifying the workflow in terms of the processing steps to be performed on a given image dataset was proposed in the requirements section. This idea was adopted in the design of ImageCompute and was implemented by a MxPlugin component developed for the TinyXML¹⁵⁹ software library with interfaces that serve to wrap its basic functionality in terms of creating, reading and writing XML files as well as their elements; an application of the Façade software design pattern as proposed by Gamma et al¹⁹⁶.

A collection of six Dynamic Link Libraries (DLLs) were developed to provide the functionality required for the specific types of XML files needed to support the workflow, as described in a further section. In addition DLLs were developed to provide common functionality shared by these libraries; ICProc and ICUtills. In this way the base code could be used by multiple applications (EXE) to perform the required workflow processing.

The various types of files created and used by the workflow are stored in the computer's file system in sets of predefined folders as shown in Figure 6-4. This

approach was preferred to storing the information in some form of relational database as it makes the data more accessible and avoids the overhead of refactoring the database during development, as discussed by Ambler and Sadalage¹⁴³. However, an alternative approach may be adopted in future versions of the workflow.

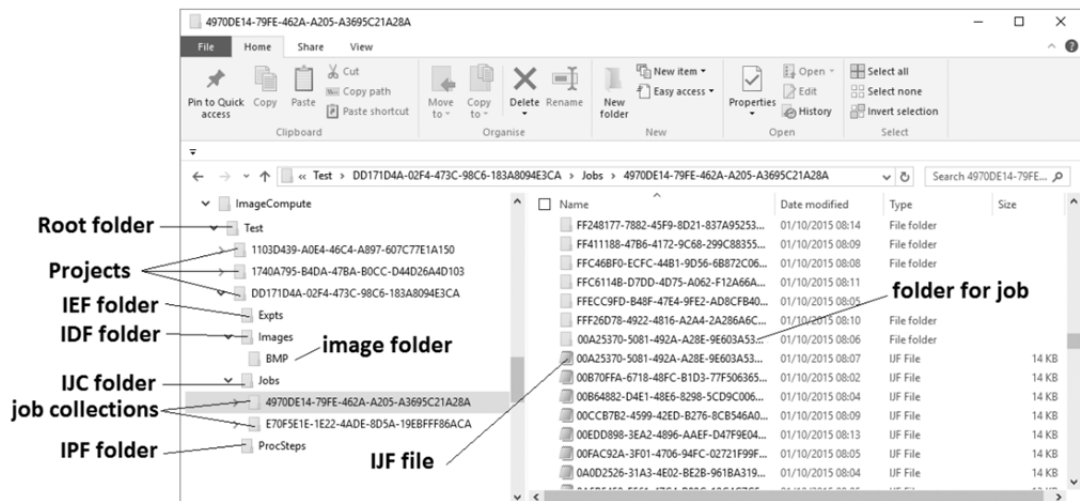


Figure 6-4: Organisation of workflow files and folders in a computer file system. The Root folder contains a collection of project folders. Each project folder contains subfolders for its experiments (IEF), the processing steps shared by experiments (IPF), image datasets (IDF) as well as the processing collections that have been run (IPC). Within each processing collection folder are IJF files corresponding to the job run for each image in the dataset. Images resulting from the execution of each processing step on each source image are contained within subfolders in the IJC folder.

Image Processing File (IPF)

The image processing file (IPF) served to define the processing steps that needed to be applied to each image in the collection; see IPC. In this sense it defines the transaction that will be executed on all images (like the PSX file envisaged during requirement analysis). The IPF file has a header containing a list of the components that are needed for the processing as well as elements such as a description of its purpose. The header also contains a list of elements called ‘variables’ which can be used to initialise the input parameters for individual processing steps, if not explicitly given. Variables can also be used to store the results of a processing step, as obtained from its output parameters. In this way the outputs of one step can form the inputs for another, or contain values that will be returned to the workflow as the result of the transaction.

The main body of the IPF file contains a list of processing steps with each step represented by elements such as:

```
<Call Seq="0" FnLabel="ProcessImage" PluginFile="PluginICPukctocs.dll" >
  <Inputs RootFolder="@RootFolder" D1mmL="@SD1mmL" D2mmL="@SD2mmL"
    D3mmL="@SD3mmL" MatchTol="0.6" />
  <Outputs D1D2Angle="@D1D2Angle" CalLnCount="@CalLnCnt"
    ImageTypeName="@ImageType" PixelsCmLS="@ScaleLS" />
</Call>
```

Code Fragment 6-4: Fragment from an IPF file forming an image processing step. In this case the ‘ProcessImage’ function is described in terms of the component that implements it (PluginICPukctocs.dll) and its parameters. It can be seen that the function needs various input parameters such as MatchTol and generates output parameters such D1D2Angle. The values of parameters can be given implicitly like MatchTol="0.6" or using a variable like D1mmL="@SD1mmL".

The IPF file is used by the workflow to generate a specific IPJ for each image found in the dataset before it starts processing the image transactions. Although it is possible to create an IPF file anew using a simple text editor, a better approach is to start from an Image Processing Sample (IPS) file created by one of the MxPluginForImages components. The IPS lists all the functions (algorithms) implemented by the component together with sets of sample input and output parameters. It can be generated by the MxPluginManagerGUI by using the ‘Export IPS’ button which appears when the IPluginImageComputeLib interface is selected; see Figure 6-3.

Image Processing Job (IJF)

The Image Processing Job (IJF) file is generated automatically by the workflow software and contains the information relating to a specific image processing transaction. It contains details about the image to be processed from the IDF and the processing steps of the transaction as obtained from the IPF. Therefore the IJF has a similar structure to the IPF, but has the particular values for the specific image transaction. The IPJ is the file processed by the workflow engine during an image transaction and therefore is considered atomic, consistent, isolated and durable as it is updated with the results only after the successful completion of the transaction, has values within ranges specified by the IPF, is independent from any other image transaction and is stored in the computer’s file system; see Figure 6-4.

The workflow creates a subdirectory for the image transaction during the execution of the image processing job. This subdirectory is used to store versions of the source image file after each processing step in order to facilitate debugging of the transactions during their development, though this feature can be disabled when the workflow is used from production purposes.

Image Processing Collection (IPC)

The Image Processing Collection (IPC) file contains the collection of image processing jobs for the execution of a given workflow engine; see Figure 6-5. It is generated automatically by the workflow software during the preparation of jobs for processing and is updated as each job (transaction) is completed to record its status and runtime error count as well as the time the job was run and its duration. Future enhancement of the workflow may allow an IPC to be split amongst a collection of workflow engines so they can be processed in parallel.

Job Identifier	Job Status	Image File Processed	Start Time	Run Time	Errors
Job JobID="076AF77-000A-4768-9000-EC900AA5514"	Status="Done"	ImageRef="1.2.826.999999.1.9900.1.20100706192927427.1.1.0"	Run="2015-10-01,09:23:28"	Runtime="0.008"	RunError="none" />
Job JobID="0807A06-4C9F-4918-8975-3813024855E"	Status="Done"	ImageRef="1.2.826.999999.1.9900.1.20100706192927427.1.2.0"	Run="2015-10-01,09:23:28"	Runtime="0.008"	RunError="none" />
Job JobID="DEC3647-583F-4284-8EF0-79951E09012"	Status="Done"	ImageRef="1.2.826.999999.1.9900.1.20100706192927427.1.4.0"	Run="2015-10-01,09:23:28"	Runtime="0.008"	RunError="none" />
Job JobID="C195A00-203C-4F82-9E8E-878999403A00"	Status="Done"	ImageRef="1.2.826.999999.1.9900.1.20100706192927427.1.5.0"	Run="2015-10-01,09:23:28"	Runtime="0.009"	RunError="none" />
Job JobID="70F8E923-6E15-4747-806E-208F2F7C893"	Status="Done"	ImageRef="1.2.826.999999.1.9900.1.20100706192927427.1.1.1"	Run="2015-10-01,09:23:28"	Runtime="0.008"	RunError="none" />
Job JobID="18C604C-3400-4278-A8C-899284C68189"	Status="Done"	ImageRef="1.2.826.999999.1.9900.1.20100706192927427.1.2.1"	Run="2015-10-01,09:23:28"	Runtime="0.008"	RunError="none" />
Job JobID="C386879D-96A3-40A1-4A73-46230674F629"	Status="Done"	ImageRef="1.2.826.999999.1.9900.1.20100706192927427.1.3.1"	Run="2015-10-01,09:23:28"	Runtime="0.008"	RunError="none" />
Job JobID="C0C2A938-114C-409C-8E71-8F41118A6A92"	Status="Done"	ImageRef="1.2.826.999999.1.9999.1.20080515104437139.1.1.0"	Run="2015-10-01,09:23:28"	Runtime="0.074"	RunError="none" />
Job JobID="D88789F-036E-4644-8918-809793020649"	Status="Done"	ImageRef="1.2.826.999999.1.9999.1.20080515104437139.1.2.0"	Run="2015-10-01,09:23:28"	Runtime="0.08"	RunError="none" />
Job JobID="787A0D5-5F1C-4318-9A4F-F81E86811FA7"	Status="Done"	ImageRef="1.2.826.999999.1.9999.1.20080515104437139.1.3.0"	Run="2015-10-01,09:23:29"	Runtime="0.071"	RunError="none" />
Job JobID="957F0913-573C-463C-8E70-C0279E03A015"	Status="Done"	ImageRef="1.2.826.999999.1.9999.1.20080515104437139.1.4.0"	Run="2015-10-01,09:23:29"	Runtime="0.081"	RunError="none" />
Job JobID="86E4890C-EF48-41FE-899A-1EF60E7587A2"	Status="Done"	ImageRef="1.2.826.999999.1.9999.1.20080515104437139.1.1.0"	Run="2015-10-01,09:23:29"	Runtime="0.102"	RunError="none" />
Job JobID="54C15380-5C8B-487D-966C-FEE667780C9"	Status="Done"	ImageRef="1.2.826.999999.1.9999.1.20080515104437139.1.2.0"	Run="2015-10-01,09:23:30"	Runtime="0.089"	RunError="none" />
Job JobID="4EE83380-902B-4818-9857-86472C193189"	Status="Done"	ImageRef="1.2.826.999999.1.9999.1.20080515104437139.1.3.0"	Run="2015-10-01,09:23:30"	Runtime="0.078"	RunError="none" />
Job JobID="8035592C-6150-4F67-8FE0-F7589825777"	Status="Done"	ImageRef="1.2.826.999999.1.9999.1.20080515104437139.1.4.0"	Run="2015-10-01,09:23:30"	Runtime="0.079"	RunError="none" />
Job JobID="24164290-0923-4C16-8480-53FE3049188A"	Status="Done"	ImageRef="1.2.826.999999.1.9999.1.20080515104437139.1.5.0"	Run="2015-10-01,09:23:30"	Runtime="0.075"	RunError="none" />
Job JobID="C4058400-2178-4F21-8130-9773AC800200"	Status="Done"	ImageRef="1.2.826.999999.1.9999.1.20080515104437139.1.6.0"	Run="2015-10-01,09:23:30"	Runtime="0.090"	RunError="none" />
Job JobID="1E02E08E-909C-44CD-834B-16A09C83A607"	Status="Done"	ImageRef="1.2.826.999999.1.9999.1.20080602100922349.1.1.0"	Run="2015-10-01,09:23:31"	Runtime="0.084"	RunError="none" />

Figure 6-5: Structure of Image Processing Collection (IPC) File. The header of the file defines its associated project and experiment. The body lists the image processing jobs (IJF) performed by this IPC to complete an experiment. The IPC contains one job for each image in the experiment's dataset (IDF).

Image Dataset File (IDF)

The Image Dataset File (IDF) contains the list of images from which a dataset is composed as well the definition of their properties. A further description of the file's contents is given in Chapter 5. It is anticipated that IDF will be created by tools concerned with the curation of image datasets, also as described in Chapter 5. The workflow software uses the IDF and the IPF in order to create an Image Experiment File.

Image Experiment File (IEF)

The Image Experiment File (IEF) contains the results of processing a given IDF using a specified IPF, so contains elements of both files. It is automatically generated by the workflow software during the preparation of a processing run and is used to create the job collection file (IJC) and associated job files (IJF). After the workflow engine(s) has completed the processing of the IJC the workflow software updates the IEF with the results of each transaction from the IJF files. This allows the image processing results to be collected together for the purposes of analysis and thus supports the further development of the workflow in terms of splitting of the IJC amongst multiple workflow engines for the purposes of parallel processing.

Image Project File (PEF)

The Image Project File (PEF) provides a way of collecting related image experiments (IEF) which typically share subsets of a common collection of images and variations of processing instructions (IPF). However, during the thesis project it was used only as a placeholder in this respect.

Design of the Workflow's Command-line User Interface - ICApps

A number of simple programs (.exe) were designed to provide a command line interface for the DLLs that provided the core of the image workflow software as described above. These were standard Windows console applications which could be executed from the operating system's command line. This allowed the workflow to be run from a series of batch files with the standard output redirected to a log file so that the parameters used for the various research experiments could be recorded as illustrated in Figure 6-6.

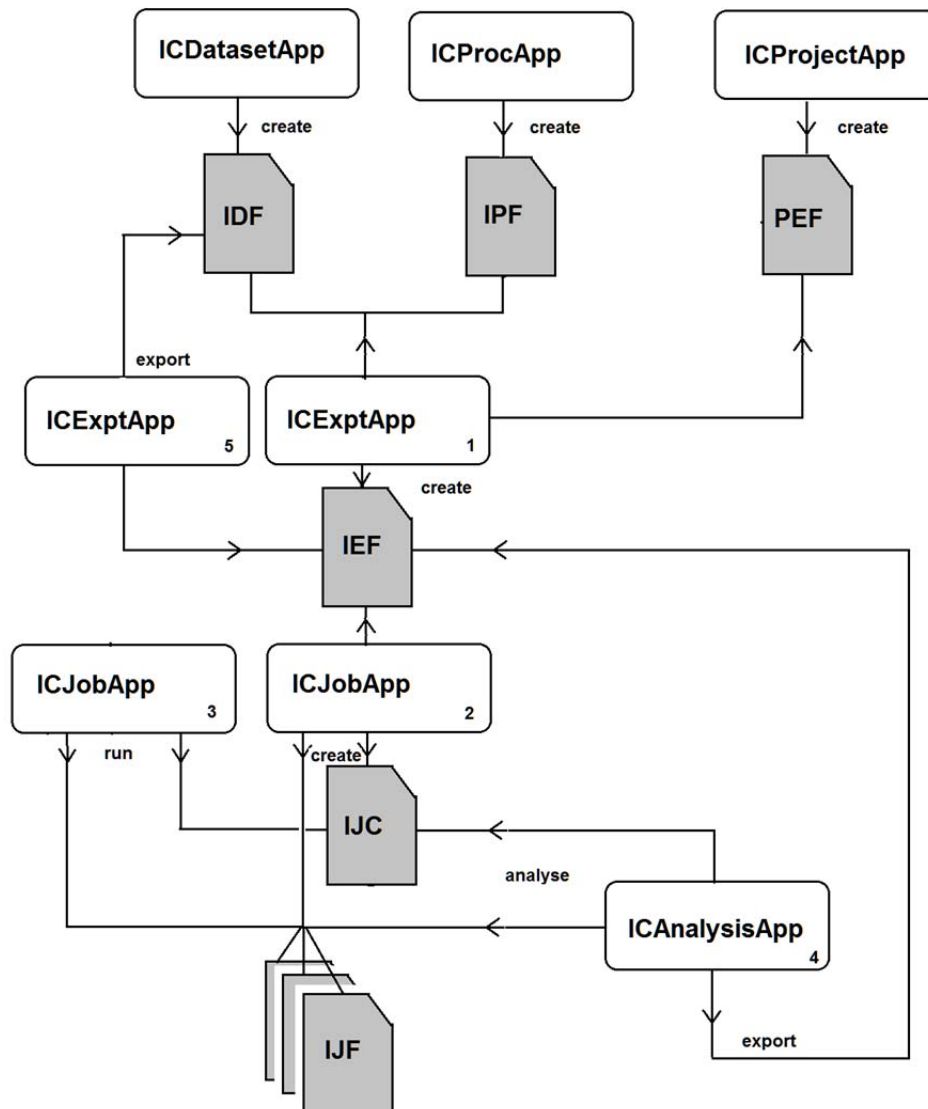


Figure 6-6: ICApps System Overview. The image workflow is implemented as a series of command line applications which operate on various document types. The application ICProjectApp creates a PEF document which acts as a container for a collection of experiments, each represented by an IEF document. This is created by ICExptApp from a dataset document (IDF) and a processing step document (IPF); see stage 1. The ICJobApp program uses the IEF to create a job collection document (IJC) as well as a collection of job documents (IJF) each of which define the processing to be performed on a particular image in the IDF; see stage 2. The ICJobApp program then runs the jobs (see stage 3) and upon completion ICAnalysisApp exports the resulting data to the IEF; see stage 4. This data may then be exported as new properties associated with the images and exams in the IDF; see stage 5.

The various command line applications were developed to prepare and run the image workflow as well as to analyze its results. They are listed as follows in order of typical execution for the completion of a research experiment:

1. **ICProjectApp** – created the project file (PEF), its directory and all its subdirectories; see Figure 6-4.
2. **ICDatasetApp** – created an image dataset file (IDF) and allowed the importation of a list of images and their properties from files generated by an application capable of accessing the contents of the URA (see Chapter 4) and the Trial Management System.
3. **ICProcApp** – created an image processing file (IPF) which could be edited using a standard text editor (Notepad) to define the processing steps for an image transaction
4. **ICExptApp** – uses the image dataset file and image processing file to create an image experiment file (IEF) containing a list of images and a mapping to associate their properties with variables to be passed to the workflow engine during the processing of a transaction
5. **ICJobApp** – uses the image experiment file to create the separate image job files (IJF) as well as the job collection file (IJC) for running the image workflow. ICJobApp was also responsible for running the workflow for a specified collection of IJFs as given by the IJC.
6. **ICAnalysisApp** – reads the results of the separate image transactions as stored in IJF files and aggregates them into the image experiment file (IEF). The IEF is essentially an XML file so it can be exported to applications such as Microsoft Excel for futures analysis of the experimental result
7. **ICExptApp** – exports properties from the image experiment file (IEF) into the image dataset file (IDF) in order to enrich the data associated with each image. For example, after running an experiment the ordering of D1, D2, D3 could be corrected for all the images in the dataset.

Although the command line interface was an effective way of running the workflow during development, it was found to be less effective when used in production for the purposes of performing research. In particular it was difficult for research

supervisors to follow the course of an experiment during the review process. For these reasons an alternative user interface was designed as described below.

Domain Specific Language Design

A Domain Specific Language (DLS) is defined by Fowler¹⁹⁷ as “a computer language of limited expressiveness focused on a particular domain”. He explains a computer language as something that is easy for humans to understand whilst still being executable by a computer and describes limited expressiveness as lacking the sort of data, control and abstraction structures of a general-purpose programming language. Therefore a DSL can be considered as a simple mechanism for someone to program a computer to solve problems related to a certain type of application. In this sense Microsoft Excel might be considered as a DSL, though it doesn't meet the criterion of being limited to a particular application.

A better example of a DSL is the simple graphical representation provided by Microsoft Access which allows non-technical people to design databases. This introduces two further properties of a DSL considered significant by Fowler; a) the possibility of the language being used by people who don't consider themselves computer programmers and b) expressing the language in graphical terms rather than as lines of text. This creates the possibility of experts in domains other than computer programming being able to use a computer to solve their problems simply by expressing them diagrammatically which can be considered as a form of language workbench¹⁹⁸.

Clearly, the implementation of a DLS for specifying the instructions that form an image transaction has potential in terms of allowing researchers to process collections of images without mastering computer programming, so moving towards the type of image processing environment proposed by Figure 6-1. The work required to create such a DSL is greatly simplified because the instructions required by the workflow software are already expressed in XML; see Image Processing File (IPF). Therefore there is no need to build a parser or the sort of semantic model it might populate in order to generate computer code, as proposed by Fowler. Instead the effort was focused on developing an editor capable of representing image processing algorithms in a graphical way and assembling them into an image transaction in the form of an IPF.

Design of the Workflow's Graphical User Interface - ImageExpter

The design of the DLS editor followed the general development approach of a Microsoft Framework Class (MFC) application implementing a multiple document interface (MDI). This application was called ImageExpter and was extended to provide a complete graphical user interface (GUI) for the workflow software tools. Its drawing features were implemented using the BCGSuite library¹⁹⁹ which allowed elements to be dragged from a panel modelling a toolbox onto the drawing surface creating the necessary links between them as shown in Figure 6-7. In this way the MxPluginForImages components found on the computer were each assigned to a different pane and the functions (algorithms) they implemented listed within such panes.

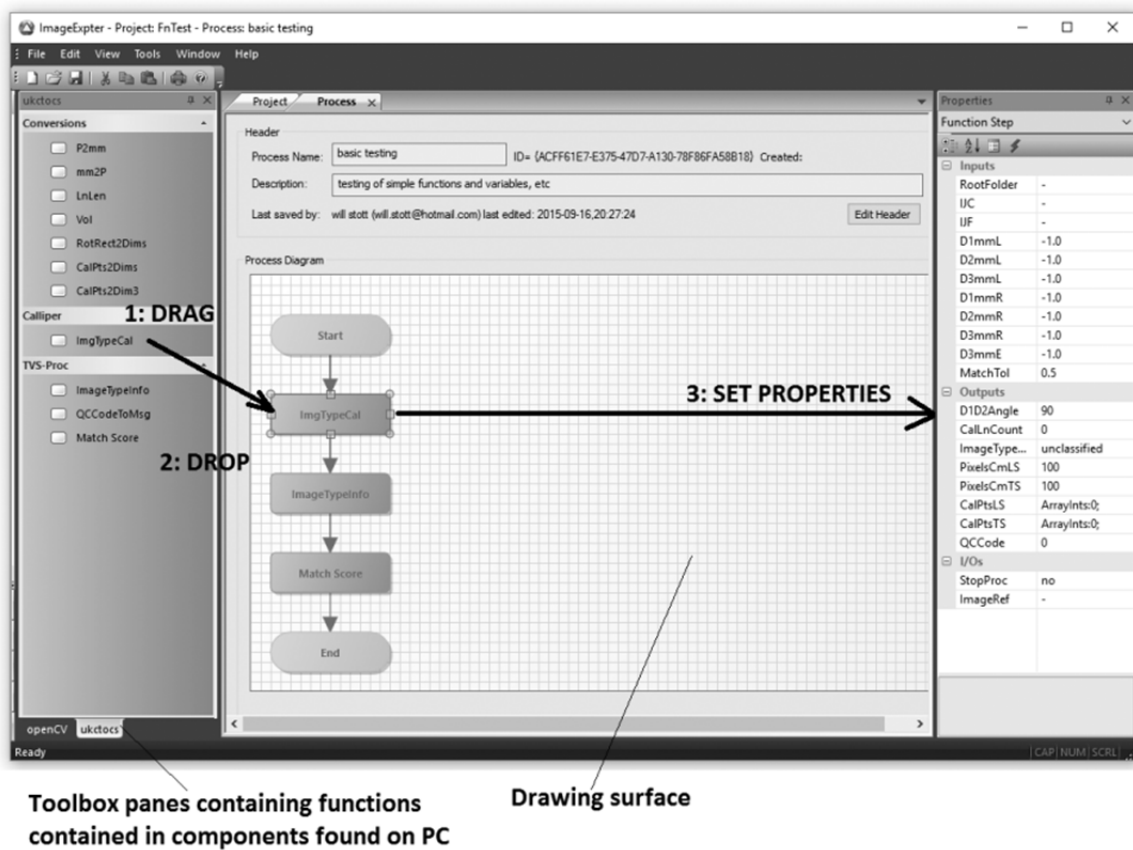


Figure 6-7: Language Workbench for creation of IPF files. The researcher creates the steps needed for an image transaction by dragging functions from the toolbox (left) and dropping them on the drawing surface (centre). After a sequence of processing steps has been created, each step can be selected so its parameters can be set using the property window (right). In this way the researcher can define the image processing for an experiment without needing to write computer code, or having to edit an IPF file directly with an XML editor.

The self-documenting feature of the `MxPluginForImages` components allowed the parameters of the algorithm to be displayed as properties and their values to be set using an appropriate user interface control. The drawing was serialised as an IPF file so it could be used by the workflow software or re-opened for subsequent editing, perhaps to create a new IPF file for use in a new image experiment.

In addition to its function as a DSL, `ImageExpt` allows the creation of image projects and the associated IEFs from IDFs and IPFs as previously described. In this way it provided an alternative to the use of command line programs for preparing and running the workflow software. However, the work required to implement these two forms of interface was significantly reduced because the workflow's core software was implemented as Dynamic Link Libraries so could be easily shared between the command line and GUI versions of the workflow application.

Design of Components for Processing UKCTOCS Images

The architecture of the image workflow software was designed to facilitate the use of a wide variety of algorithms, not only those developed specifically for a particular project but also those with more general application obtained from external sources such as OpenCV. In this later regard the approach of developing an `MxImage` component that wraps an existing C++ library and makes it accessible to the workflow by providing the necessary interfaces has already been described in terms of the `TinyXML` library. The same method was used to wrap the OpenCV library, though the interfaces provided were those of a `MxPluginForImages` component. In addition, a separate component was developed for the custom algorithms needed to process UKCTOCS images.

Further algorithms could be made available to the image workflow in the future either by extending these components, or developing additional `MxPluginForImages` components. The loose coupling between such components and the workflow software is typical of a well-designed object-oriented system as it makes the system open for extension, but closed for modification; the open-closed design principle previously described. The work required to develop such components is substantially less than would be required had the algorithms been embedded inside the workflow software, particularly when retesting and redeployment are considered. It is also

work that can be readily packaged-up and given to a third-party software subcontractor such that the resulting product can be integrated into the workflow simply by running its installation program.

MxPluginForImages component - OpenCV

The MxPluginForImages component for the OpenCV library provided the interfaces necessary for some of its algorithms to be made available to the image workflow so they could be used as part of an image transaction. The work necessary to expose an OpenCV algorithm in this way wasn't significant and typically was completed within a few hours, though additional time was needed for setup and testing. Details of the algorithms made available for the thesis research studies are given in Chapter 7 and 9. The installation program for the OpenCV component setup its files on the target computer and also performed other tasks, including the installation of the OpenCV product if required. Therefore at little more than the click of a button, the MxPluginForImages component and all its dependencies, including the OpenCV library, could be integrated with the workflow software on the target computer.

A freelance developer was engaged through the website 'People per hour' in order to test the viability of using a third-party software contractor to implement algorithms in an MxPluginForImages component, as proposed in Figure 6-1. The developer worked remotely (India) and managed to successfully add 18 OpenCV algorithms to the component in approximately 120 hours. This time included an allowance for him to familiarise himself with the architecture, recreate the development environment, write a full suite of unit tests and deliver the solution including its installation program. He estimated it would take considerably less time to add further functions as much of this work was preliminary setup tasks that would not need to be repeated.

MxPluginForImages components – UKCTOCS

In addition to the type of general purpose image processing algorithms provided by OpenCV the thesis project required the implementation of some algorithms specific to the processing of UKCTOCS images. For example, an algorithm was needed to locate the calliper marks in a TVS image. These specific algorithms were implemented in a separate MxPluginForImages component which provided the interfaces necessary for them to be invoked by the workflow software as part of an

image transaction. Details of the algorithms needed for the thesis research studies are given in Chapter 7 and 9.

Discussion of use of the workflow in research

The image workflow was implemented according to the design which has been summarised in the previous sections. The first release of the core workflow software and command line interface was delivered in Q3 2014 and was used for the research described in Chapters 7 and 9. The graphical user interface was completed in Q3 2015 and will be enhanced during 2016 with the aim of releasing a product that will allow research on the images in the URA by people who do not have experience in computer programming as proposed in Figure 6-1. These applications of the image workflow are summarised below.

Implementing Quality Control for UKCTOCS Images

The image workflow is well suited to automating the collection of quality control (QC) information in respect of the UKCTOCS ultrasound images as it can measure a variety of metrics for a large number of images. Although during the thesis project it was used to measure only a few simple QC metrics for a relatively small dataset of approximately 4,000 images, there is obvious potential to generate a wide number of complex metrics for much large datasets. For example, the prototype work described in Chapter 9 involving the detection of ovary boundaries using an active contour model could be extended with machine learning techniques to build a classifier able to discriminate more reliably between images containing calliper marks that truly identify an ovary and images that do not. This would build on the work already completed to allow the creation of datasets of TVS exams with unsatisfactory exams excluded on the basis of an objective and independent retrospective QC rather than on the basis of sonographer self-assessment at the time of examination. Such datasets would clearly be more useful for research seeking to identify associations between such things as small changes in ovarian volume and the later onset of ovarian cancer as a significant amount of poor data would have been omitted.

Future Work

The image workflow has potential for use in order to recover information that was not recorded at the time of scanning for many more exams with images stored in the URA, such as the exact image used to measure the ovary and the ordering of the D1, D2 and D3. This would add to the value of the data in the URA which has been collected at great expense and effort, so making it more useful for the purposes of further research.

In addition, future medical trials involving the large-scale creation of images, like those generated by UKCTOCS, might benefit from further adaption of the workflow software to automate their QC metric gathering. This could permit such metrics to be used for the purposes of driving improvement in the image creation process, for example by identifying examiners who were consistently making poor images or were making mistakes like measuring certain dimension in transverse rather than longitudinal section. In this way the quality of the data collected by the trial could be improved as it progressed.

Further Development

Although the workflow's command line interface was used for the research described in Chapters 7 and 9, the graphical user interface provided by ImageExpter has clear advantages for use by people who are not computer programmers. The focus of this work will be enhancing the software so that researchers can perform all the tasks necessary to create and run an image workflow for their experiments (and archive them) so that the command line interface can be retired. In this way it is envisaged that ImageExpter will become a type of Integrated Development Environment (IDE) for image processing, in much the same way that Visual Studio is considered an IDE for software development. Work is also planned to provide better integration between ImageCurator (see Chapter 5) and ImageExpter so that image datasets can be easily prepared and annotated by clinicians and then used by researcher in their experiments.

One of the key aims for ImageExpter is that it should provide a more convenient environment for conducting image research, one that facilitates the rapid iteration of

experiments necessary to select the best algorithms and optimise their parameters. This process is already greatly assisted by the availability of images showing the results of each step of the image transaction; see Figure 6-8. However, further development will make improvements to facilitate not just the visualization of images after a processing run, but also to help with the development of the image transaction processing instructions (IPF) before processing starts. In this regard a debugging feature is planned to allow the execution of individual steps of an image transaction on a selected image directly from the DSL language workbench; see Figure 6-7. This will allow rapid assessment of changes to algorithms and their parameters in terms of their impact on the selected image.

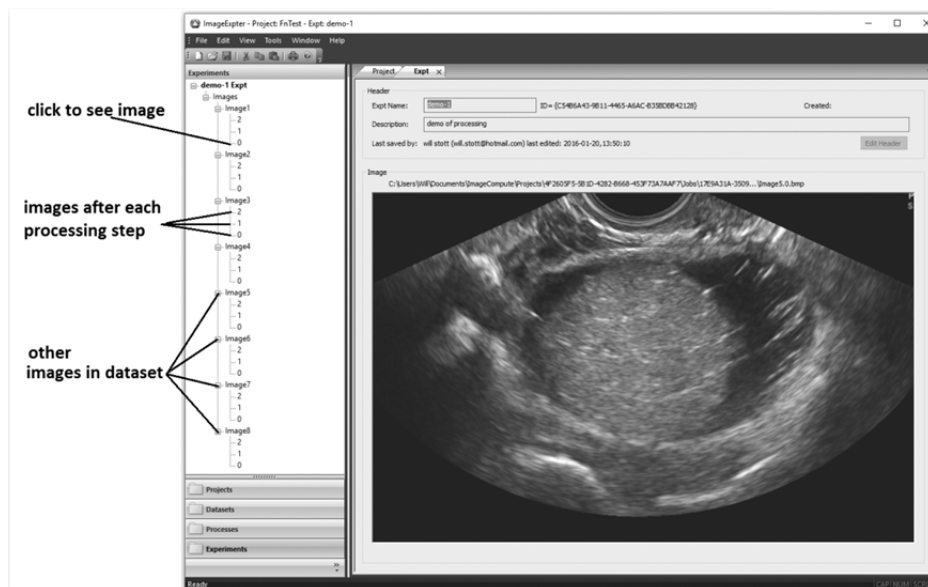


Figure 6-8: Optimisation of algorithm and parameter selection can be achieved by selecting an image and then stepping through the processing steps defined in the IPF much like a software debugger stepping through code statements. This would allow the impact of parameter or algorithm changes to be viewed almost instantaneously.

In addition to work on the enhancement of ImageExpt effort needs to be given to the development of the MxPlugin open-source project in order to stimulate interest amongst the people who develop image processing algorithms. This will allow further progress towards the type of image processing community presented in Figure 6-1 whereby groups of people with different specialisms can work together in order to solve difficult image processing problems. This includes analysis of TVS images to allow the gathering of more sophisticated QC metrics as well as many other problems in medical and other fields.

Summary

The image workflow tool presented in this chapter is capable of performing large-scale processing of images and might be extended in order to provide routine analysis of images from TVS examinations stored in the type of PACS archiving systems commonly used in NHS hospitals. Therefore it could be used to regularly gather QC metrics from such examinations performed as part of any future national screening programme for ovarian cancer. In addition, it could be used to recover information lost during the UKCTOCS trial such as ovary dimension ordering and images used to measure ovaries, as described in Chapters 7 and 9 of the research part of the thesis.

The workflow tool also has potential for other types of image processing not just in terms of ultrasound, but other types of medical and non-medical images as well. Furthermore, as the tool can be used by people who lack computer programming skills, it might greatly reduce the barriers to entry in the field of image research and encourage people to engage in this work who might currently be put-off by the technical challenges. In this respect it can be hoped that it will encourage people with specialist skills to form cooperative teams engaged in tackling the wide range of image processing challenges faced in modern medicine and other industry sectors.

Research Chapters

Chapter 7: Using the Image Workflow for TVS Quality Control

Introduction

Visualization Rate (VR) is the key Quality Control (QC) metric for TVS examinations. It can be defined, as stated in Chapter 3, as the number of exams that resulted in ovarian visualization divided by the total number of exams performed over a given period. Typically VR assessment is based on the report of the examining sonographer as to whether or not the ovaries were seen for each of her TVS scans. However, this is suboptimal in terms of QC as the metric is self-assessed.

It is rare to conduct an independent validation of ovarian visualization at the time of examination, as it requires a second professional either to observe or repeat the scan. Therefore such validation usually involves a post-examination review of the static images made to measure the ovary. This also requires significant resources, particularly if experts in gynaecological scanning are involved in the review. Consequently, it is the aim of this thesis to show that software tools may help in the collection of independent and objective metrics for TVS exams by automating the analysis of the static images generated during scanning.

The previous chapter described the development of a workflow tool able to process the thousands of static images that would need to be analyzed if it was decided to produce QC metrics from TVS exams on a routine basis. This chapter describes the adaptation of this tool to gather basic metrics with QC potential as well as detailing a study that was undertaken to assess its capability as a measurement device. The findings of the study have been submitted for publication in *Ultrasound in Obstetrics & Gynecology*²⁰⁰. The co-authors collaborated in the design of the study and the generation of the data as well as its interpretation. All authors contributed to the editing of the paper parts of which are reproduced in this chapter.

Background and Objectives

The trial protocol required the sonographer to record the linear size of the ovary in terms of its dimensions D1, D2, D3 as well as the thickness of the endometrium; more complete details are given in Appendix B. This was achieved by saving static images from the ultrasound machine's cine video memory so calliper measurements could be made as shown in Figure 7-1. In this way the sonographer could read the dimensions displayed at the bottom of the image and then record them on the paper exam report for later transcription into the Trial Management System (TMS). These static images were saved in DICOM files and stored on the ultrasound machine's hard disk for subsequent transfer to the coordinating centre (CC) and import into the URA, as described in Chapter 1.



Figure 7-1: Example of a TSLS-RO image containing right ovary in transverse (right) and longitudinal (left) section as indicated by the calliper marks. The corresponding dimensions are shown at the bottom of the image.

The trial protocol did not ask sonographers to record the DICOM references of the images used to measure the ovaries and endometrium thickness. Therefore it was

necessary to identify which three (or five) images had been used to make these measurements from as many as twenty-six^a other images associated with the exam, for only these images were relevant for QC purposes – e.g. confirming whether or not the object measured in the image was an ovary. The task of identifying such images can be accomplished by matching the dimensions recorded in the TMS with either the numeric values displayed at the bottom of the image, or with the values obtained by measuring the calliper distances in the image. The later approach was adopted by the thesis project and involved adapting the image workflow (Chapter 6) so it could undertake the required processing of the image dataset to match TMS dimensions with the calliper dimensions found in the images. The main technical challenge of the work was optimising the number of exams whose images could be matched in this way. For the ability of the tool to reliably identify the exact images used to measure the ovary is a prerequisite for its use in most other types of quality control for TVS examinations. Therefore a key aim of the study described in this chapter was to provide evidence of the tool’s capability in this respect. The study also aimed to show that the data generated to identify such images provided useful information in terms of some basic QC metrics for TVS exams.

Methods

The study data was formed from a collection of 1,000 TVS exams with images stored in the URA which satisfied the inclusion criteria:

- a) Annual TVS exams performed after 1st January 2008.
- b) Both ovaries visualized by the sonographer.
- c) Both ovaries classified as having normal morphology.

Transabdominal only exams and those classified as abnormal or unsatisfactory were excluded. Sonographer recorded dimension values (D1, D2, D3) for left ovary (LO) and right ovary (RO) were extracted from the Trial Management System (TMS). These randomly selected exams were performed by a team of sonographers who were all at least a certified ultrasonographer or a trained midwife or doctor with

^a Analysis of the annual TVS exams performed after 1st January 2008 for which images are stored in the TMS show that each exam has a mean of 5.03 images with a range of 1-31; see Appendix A. Most of these exams measured left and right ovary as well as endometrium thickness using three separate images, but some used four or five images to make these measurements as shown in Table 7-1.

experience in gynaecological scanning, particularly TVS. In addition, they all had additional training for assessment of postmenopausal ovaries and were subject to level 1 annual accreditation by the UKCTOCS National Lead sonographer, as discussed in Chapter 3. This dataset is the same as used for the study reported in Chapter 9.

Identifying the Images used to Measure the Ovaries

Analysis of the URA database shows that a number of images have been recorded for each TVS exam (mean 5.03, range 1-31), but only some were used to measure the ovaries. Therefore the image workflow processing needed to identify these images; a task made more challenging because of the lack of information that could have been provided by the ultrasound machine in the DICOM header.

Recovery of Data Missing from the DICOM Header

The data missing from the header can be recovered by analysis of the image in order to obtain the distances between corresponding calliper marks forming a line. These distances may then be compared with values stored in the Trial Management System (TMS) for the dimensions D1, D2, and D3. In this way it can be assumed that an image with calliper line lengths matching those recorded in the TMS was the actual image used by the sonographer to make the measurements. However, this approach presents a number of possible problems:

- **Image not recorded or error** – the image used by the sonographer to measure one or both ovaries may not have been stored in the URA. Therefore the collection of images for the exam is incomplete and no match can be found. Alternatively, the image used to measure the ovary contains some form of flaw which prevents its analysis.
- **Unresolvable** – the ovary dimensions in the TMS are the same for both left and right ovary, so the exact images used to make the measurements cannot be differentiated.
- **Duplicates found** – the sonographer recorded two or more images with the same calliper lengths but in different locations, so the exact image used to make the measurement cannot be differentiated.

In addition to the above problems, it was discovered during prototyping work that the values in the fields of the TMS record for left and right ovary dimensions did not correspond to the standard ordering of D1, D2 and D3. For example, the image might show the values D1=15mm, D2=9mm, D3=8mm but the fields in the TMS might have recorded D1=8mm, D2=15mm, D3=9mm. This problem is attributed to the failure of the examination report form to explicitly identify these fields separately; see Appendix C. Accordingly it was anticipated that a large number of examinations would have these dimensions recorded in the wrong TMS field. For this reason, the dimensions obtained from the images need to be matched to values recorded in the TMS without regard to their order.

A further challenge to the matching of ovary dimensions obtained from images with those recorded in the TMS was the discovery during prototyping that the ultrasound machines have a bug in the software used to calculate length from the positions of two callipers; see Figure 7-2. This results in small inaccuracies in the displayed values which are copied by the sonographer into the TMS. Consequently, the lengths calculated from the images do not exactly correspond to the values in the TMS which means that any match must be within a given tolerance.

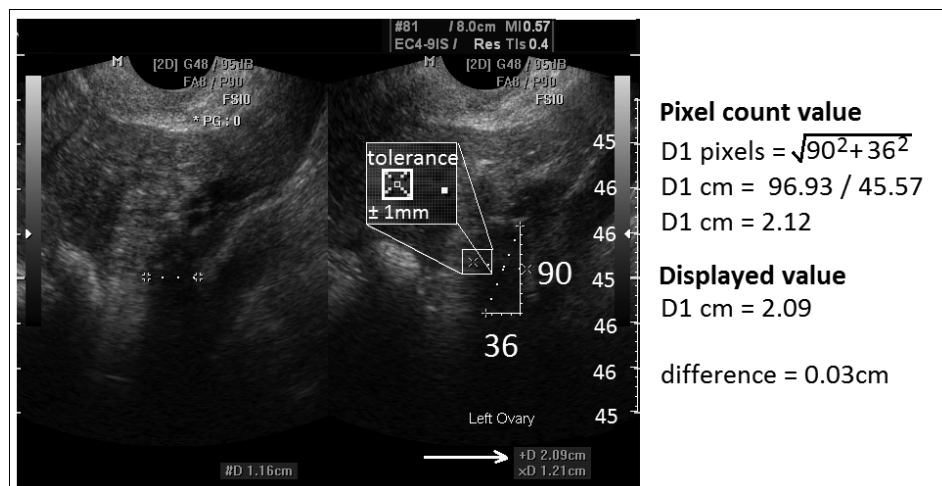


Figure 7-2: Defect in the calculation of calliper dimensions by the ultrasound machine. It can be seen that value of D1 is displayed at the bottom of the image as 2.09cm, but this value doesn't correspond to the distance between the corresponding calliper marks calculated as 96.93 pixels, or 2.12cm when the scale of 45.57 pixels per cm is applied. This small difference is irrelevant for the purposes of clinical judgement as the tolerance of such measurements is given as ±1mm (see Chapter 2). However, the difference is important when attempting to match the length of calliper lines found in the image with values recorded in the TMS as copied by the sonographer from the displayed value.

Image Formats

When measuring ovary dimensions, two views are needed; one shows the ovary in longitudinal section (LS) and the other shows it in transverse section (TS). Usually such views are contained in a single image file like that shown in Figure 7-1. The sonographer measures the size of the ovary by placing calliper marks for D1 on the longest line that can bisect the ovary boundary in any view of the ovary. This identifies the LS view. The longest line at approximately 90 degrees to D1 that can also bisect the ovary boundary in the LS view is marked by the D2 calliper. The longest line that can bisect the ovary boundary in the other view of the ovary is marked by the D3 calliper. This identifies the TS view. The ultrasound machine displays the distance between the various calliper marks in millimetres as shown at the bottom of Figure 7-1.

Although the UKCTOCS instructions specified that images used to measure ovaries should be recorded in the split view format containing both longitudinal section (LS) and transverse section (TS) in the same image file, this requirement was not enforced or subject to quality control. Therefore it was thought possible that a number of TVS exams would contain images of ovary LS and TS in separate files and this was confirmed during early prototyping. In order to help assess the extent of this problem the image workflow was required to identify the format of the images as shown in Table 7-1. In addition, formats were defined for Mismatches (image with 1, 2 or 3 calliper lines with lengths not matching the TMS values for ovary or endometrium values) and Unclassified (no calliper lines or more than three found). The left and right ovary formats are assigned by matching line lengths in the image against values in the TMS without regard for their order, as previously mentioned.

The task of identifying the formats of images was further complicated by the discovery during early prototype work that some sonographers routinely measured D3 in the LS and then measured D1 and D2 in the TS; again UKCTOCS instructions were not being followed and the problem had not been addressed by quality control. This complicated the task of achieving standardized measurements. In cases of TVS exams with D1 and D2 measured in TS the image workflow defined any image with two crossing calliper lines as LS, with D1 formed by the longer line and D2 formed by the shorter line. A non-crossing calliper line was then defined as measuring the

TS of an ovary (D3), even if it was the longest line. For an image with only one such calliper line, it would be classified either as TS-LO, TS-RO or Endometrium depending on the match with TMS values.

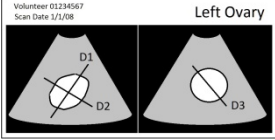
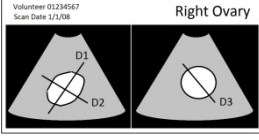
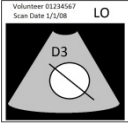
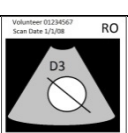
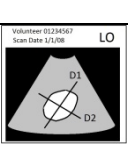
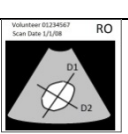
Image Name	Image Type	Description
TSLs-LO		<p>Left ovary in TS, LS in split view</p> <p>3 calliper lines (2 crossing; $D1 > D2$, $D3$ non-crossing)</p>
TSLs-RO		<p>Right ovary in TS, LS in split view</p> <p>3 calliper lines (2 crossing; $D1 > D2$, $D3$ non-crossing)</p>
TS-LO		<p>Left ovary in transverse section (TS) in single view</p> <p>1 calliper line ($D3$)</p>
TS-RO		<p>Right ovary in transverse section (TS) in single view</p> <p>1 calliper line ($D3$)</p>
LS-LO		<p>Left ovary in longitudinal section (LS) in single view</p> <p>2 calliper lines (crossing; $D1 > D2$)</p>
LS-RO		<p>Right ovary in longitudinal section (LS) in single view</p> <p>2 calliper lines (crossing; $D1 > D2$)</p>
Endometrium		1 calliper line in one view

Table 7-1: UKTOCS TVS Static Image Formats are determined from the number of calliper lines found in the image, their relative lengths and whether or not they cross. Left and right ovary images are differentiated using values from the TMS.

Exam Formats

The workflow had to process all the images for a given TVS exam in order to obtain the dimensions D1, D2 and D3 for both the left and right ovary. These images may exist in any of the formats specified in Table 7-1 and depending upon their combinations the exam would then need to be classified as shown in Table 7-2.

The lack of standardization of image formats combined with the small errors in the values recorded in the TMS and the lack of certainty about their ordering makes the matching of dimensions obtained from the images against those in the TMS a significant challenge. Accordingly, it was anticipated that it would not always be possible to match the two sets of dimensions so as to identify the exact images used to measure the ovaries. This would be a particular problem when the dimensions recorded in the TMS had similar values, particularly if a number of images had just one calliper line.

Exam Name	Left Ovary Measurements	Right Ovary Measurements
TSLs-Both 2 images		
TSLs-Left 3 images		
TSLs-Right 3 images		
TSLs-None 4 images		

Table 7-2: UKCTOCS TVS Exam Image Formats depend upon the combination of image formats identified in Table 7-1. It should be noted that only the exam format TSLs-Both complies with UKCTOCS instructions given to sonographers.

Image Workflow Processing Requirement

The image workflow required adaptation so it might process each image in the dataset to:

- Find any calliper marks and also obtain the distance between matching pairs in millimeters using the scale marks at the edge of the image.
- Find the angle between any calliper lines that crossed each other.
- Find the format of the image as described in Table 7-1 by matching the lengths of calliper lines against values in the TMS.

The above information needed to be recorded in the Image Dataset File (IDF) in the case of images that had calliper lengths matching dimensions recorded in the TMS so subsequent processing would allow identification of exams that either had a complete set of images corresponding to one of the TVS Exam image formats (Table 7-2), or were considered incomplete due to duplicates, unresolved images or images not recorded.

Finding the Calliper Marks

During a TVS examination the sonographer uses controls on the ultrasound machine to place calliper marks on a static image, as previously described. The machine then draws a dotted line between two associated calliper marks and displays the distance between them at the bottom of the image. Slightly different styles of calliper marks are used for each separate measurement. For example, in Figure 7-1, a cross, a saltire and negated cross form the pairs of calliper marks used to measure the dimensions D1, D2 and D3 respectively.

UKCTOCS did not standardise the use of styles of calliper marks for measuring each of the different ovary dimensions, so a variety of styles are used for D1, D2 and D3 in addition to those shown in Figure 7-1. However, they are generally shown in green and dotted lines are shown in yellow, though some variation in colour was found during early prototyping work. It is easy to differentiate a calliper mark because each of its pixels share the same value, unlike the variation in grey-scale that exists between most adjacent pixels in the underlying image which arises due to ultrasound speckle (as described in Chapter 2).

Finding the Length of Calliper Lines

The image also contains scale markers on the left and right edge as shown in Figure 7-1. The pixel count between the larger scale marks represents the distance of a centimetre in the image. This scale can be used to translate lengths in pixel counts into physical distance in millimetres. Therefore the physical distance between two calliper marks can be obtained by counting the number of pixels between them in the X and Y-axis in order to calculate the length of the hypotenuse before using the scale markers to translate this length to millimetres.

The image workflow needs to locate all the calliper marks that exist in an image and then find all pairs linked by a dotted line and which also share a common style. This allows calculation of the length of the various lines in the image that were used to measure the ovary dimensions.

Finding the D1, D2 Angle

The image workflow needs to measure the angle between the lines used to measure D1 and D2. This is important as the prolate ellipsoid formula used to calculate the volume of the ovary from its dimensions D1, D2 and D3 assumes that the lines D1 and D2 are perpendicular to each other, so any error in their angle will cause an inaccuracy in the volume calculation. The angle of each line can be measured by considering it to be the hypotenuse of a triangle with the other two lines being formed by the Cartesian distances in the X and Y axis between the two points that mark its start and end location. The angle of the line with respect to the X-axis can then be calculated by standard trigonometry and thereby the difference of angle between the two lines may be determined

Finding the Format of the Image

The lengths of calliper lines found in the image needs to be matched against the dimensions recorded in the TMS in order identify whether the image may have been used by the sonographer to measure one of the ovaries. If, after processing all the images for given examination in this way only one TSLS-LO image and one TSLS-RO image are identified, then it can be deduced with certainty that these images were the ones used by the sonographer to measure the ovaries. It can also be said that the objects marked by the callipers in these images were considered to be ovaries by

the sonographer at the time of examination. Similar conclusions can also be reached in the case of examinations with images in different formats, though the processing is more convoluted.

Design of Processing Components

The processing requirements were implemented by the development of a number of algorithms packaged into a custom C++ library; MxPluginForImages. This library implemented the interfaces needed for its integration with the image workflow as described in Chapter 6. The designs of the main algorithms (functions) required by the study are described in the following sections.

Algorithm for Finding Calliper Lines and their Properties

Identifying the location of overlaid marks in an image like arrowheads or callipers would seem to be a common requirement, but few techniques reported in the literature address this problem. The approaches identified can be categorised into sparse pixel vectorization, segmentation of symbol-like objects and global or local thresholding. In addition Santosh et al²⁰¹ describe a method based on template matching via Dynamic Time Warping (DTW). These techniques are summarised as follows:

- **Sparse pixel vectorization**²⁰² (SPV) – detection of a few (sparse) selected points along an axis in order to create a vector containing positions at which certain types of pixels can be found with the aim of subsequently removing positions that are redundant in order to describe the detected object more succinctly. For example: a straight line might start with a vector containing ten locations representing various points on the line, but then be reduced to just its start and end locations. SPV was used by Dori and Wenyin²⁰³ to detect various graphical elements in a drawing. This might have application in finding the lines between callipers, but would be problematic for TVS images as the lines are already formed by sparse groups of pixels (dots).
- **Segmentation of symbol-like objects**²⁰⁴ – small objects and noise are removed from the image before it is processed to identify the existence of objects matching a mask which are then used for feature computation. This approach requires the whole image to be processed to identify pixels with

common properties before attempting to identify the regions likely to form calliper marks.

- **Global or local thresholding**²⁰⁵ – detection of pointer annotations by first detecting their edges (Roberts, Sobel, Canny, etc.) and then applying adaptive thresholding to reveal their boundaries. This approach would not be applicable to callipers as they are small blocks of pixels without a separate edge.
- **Template matching via DTW** – the image undergoes fuzzy binarisation at four different threshold levels each of which are then segmented using the connected-component principle – i.e. detect clusters of pixels with the same value connected at their edges (8) or faces(4). These clusters are matched to templates by finding similarities between the two sequences of values using the same sort of dynamic time warp algorithm used in applications such as voice recognition. This approach was intended to find large arrowheads having various shapes and orientations as well as noise.

The above approaches, though interesting, were not used to develop the function used to find the callipers in the study. This was due to concern about their applicability as well as the processing overhead they would require and its impact on the speed at which images might be processed in the workflow. The requirements for calliper detection in the UKCTOCS images are much simpler as there is little noise to consider and the calliper shapes are known in advance. Therefore the calliper finding function was based on a template matching approach summarised as:

1. Define the shape and size of all the callipers likely to be found in UKCTOCS images in terms of two-dimensional vectors containing Boolean datatypes. These form templates that can be matched against the pixels in the image such that TRUE indicates that the corresponding image pixel will be tested and FALSE indicates that it will not. When a pixel in the image is tested its value is compared to the standard calliper colour for green – i.e. red, green, blue value of 0,255,0.
2. Overlay each template in turn at the top left corner of the image and test for a match against the underlying pixel values. This might be implemented efficiently by performing a logical AND operation for each value in the two-

dimensional template vector against the corresponding pixel location in the image. The process should stop if a mismatch is detected and then continue with step 3; otherwise if each pixel corresponding to a TRUE value in the template has a value matching the calliper colour, then it can be assumed that a calliper has been found. In this case its location and type need to be recorded in a list.

3. Move the template one pixel to the right with respect to the image and repeat step 2. Repeat this operation until the end of the row is reached, then proceed with step 4.
4. Move the template one pixel down with respect to the image and repeat step 2 and 3. Repeat this operation until every row (image) has been processed, then proceed with step 5.
5. Process the list of calliper types and locations in order to find pairs of matching types. Calculate the positions of a line between each pair and then test these positions for the existence of clusters of pixels forming the dots of a line. Record any lines that are found together with their start and end locations.
6. Calculate the pixel lengths of the calliper lines found in the image and convert the values to millimetres using the scale at the edges of the image – i.e. number of pixels between markers corresponding to 10mm. Record the line lengths in millimetres.
7. Determine if any of the lines in the image cross each other by checking whether any of their points coincide. Calculate the angle between lines that cross and record the value.

Processing speed is optimised by avoiding the need for image pre-processing and the cessation of attempts to match the pixels upon the first pixel mismatch against all candidate templates. In this way after each image is loaded from its file a corresponding image object is created with information about the presence of calliper lines, their length and the angle that they cross each other (if they do).

Algorithm for Classifying Image and TVS Exam Formats

The problem of classifying the images and TVS exam format is considered specific to UKCTOCS so there is no similar work performed by others from which to learn.

However, given that the callipers in the image have been found and the number, length and angles of the associated lines have been determined (as described above), then performing such classification is simple. It is a matter comparing the properties of the image to the criteria in Table 7-1 to determine the type of image and then comparing the appropriate dimensions from the TMS to the line lengths in order to decide whether it might have been used by the sonographer to measure the ovaries. In the case of images that satisfy such checking the image type and the correctly ordered dimensions as found in the image are passed back to the workflow for storing in the Image Dataset File (IDF). When comparing the dimensions of the line lengths found in the image with those recorded in the TMS it is necessary to allow some tolerance due to the bug in the ultrasound machine previously described; see Figure 7-2. The consequence of this defect is that the value displayed in the image is always not exactly the same as the value determined by measuring the calliper lines in the image. Therefore the sonographer recorded values may have a small error.

During the testing of the functions it was noted that the number of examinations whose format could be determined in terms of Table 7-2 varied considerably according to the value used as the tolerance for matching dimensions. The results of this experimentation are shown in Figure 7-3. It suggests that in order to optimise the number of exams that could be categorised by the software the tolerance of matching dimensions stored in the TMS against those obtained from image analysis should be set at $\pm 1.0\text{mm}$.

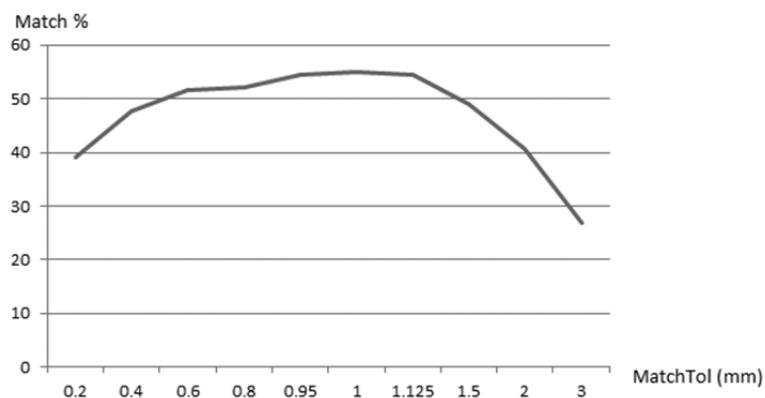


Figure 7-3: Optimization of matching tolerance between TMS and image dimensions. The success of identifying the images used to measure the ovary was dependent on matching within a given tolerance the length of calliper lines found in image against the dimensions recorded in the TMS. Experiments were made using a number of different tolerances (MatchTol) and a tolerance of 1.0mm was found to be optimum.

After each image in the dataset has been processed the IDF can then be analyzed by its examination elements in order to set the examination format property as defined in Table 7-2. However, such property values are only set for examinations that have a unique image for the various dimensions; otherwise examination format is set as duplicate, unresolvable, image not found or processing error.

Implementation of Workflow Processing

The first phase of the study was concerned with validating the capability of the system to accurately collect the above data. This involved:

1. Using the software to identify the images used to measure dimensions by matching values recorded in the TMS with those found in the image.
2. Measuring by hand the dimensions in the images from a sample of exams to validate the values measured by the software.
3. Repeating the processing performed by the software to confirm that the same dimension values were produced in each run.
4. Validating the software in terms of its ability to measure dimensions by comparing ovary volumes calculated from dimensions recorded in the TMS against volumes calculated from software measured dimensions.

The second phase of the study was concerned with calculating the QC metrics using data collected from all exams with images identified and analyzing them statistically to show that the system has potential for use in a clinical setting.

Statistical Analysis

The study's statistical analysis involved examination of the image dataset file (IDF) upon completion the workflow's processing run. This was facilitated by importing the data into a spreadsheet.

Identification of Images used to Measure Ovaries

The images were classified by format as described in Table 7-1, with two further classifications applied:

- 1) **Unresolved** – images that could not be identified because the length of their calliper lines did not uniquely match the measurements stored in the TMS for

the exam. For example, a TS image with calliper length 8mm cannot be resolved if the TMS values for LO and RO both contain a dimension of 8mm.

- 2) **Invalid** – images with callipers or scale marks that could not be processed; unknown calliper types, scale marks not found in expected locations, odd number of calliper marks found, etc.

The exams were then classified by their image formats as TSLS-Both, TSLS-Left, TSLS-Right and TSLS-None (see Table 7-2) with additional categories for:

- 1) **Image not recorded or error** – exams that did not have a complete set of images for LO, RO and uterus because they were not recorded in the URA, or the required images were classed as invalid as described above.
- 2) **Exact image not identified** – divided between exams with images that were:
 - a. Unresolved as defined above.
 - b. Duplicates as more than one matching image of the required format type was found. For example, the exam with 2 x TSLS images for LO that both match the dimensions for the LO in the TMS (within ± 1 mm).

The number of exams in each of the above formats was counted and the results displayed in a pie chart (Figure 7-4) with stratification between exams with images identified and exam with images not identified.

Repeatability of Results

The workflow software was used to process all the images for the exams in the study dataset on two occasions. The results of each run were compared to allow assessment of the system's ability to generate repeatable results

Manual Review

The correctness of the workflow software was validated by performing a manual review of its processing using a random selection of exams taken from all the exams identified by the software as having images matching the exam formats TSLS-Both, TSLS-Left, TSLS-Right or TSLS-None; see Table 7-2. The data recorded by the workflow software for these exams was compared with the data measured by hand in order to validate 1) the image format categorical variable and the exam format

categorical variable as specified in Tables 7-1 and 7-2; 2) ordering and length of the D1, D2, D3 dimensions; 3) the angle between the D1 and D2 calliper lines for the LS images.

In addition, a random selection was taken from all the exams that could not be identified by the software. The image format categorical variables recorded by the software for the images in these exams were compared with the values determined by manual inspection.

Agreement of Ovary Volumes: Image vs TMS Values

Further validation of the correctness of the workflow software processing in respect of measuring D1, D2, D3 dimensions was performed using all the exams identified by the software as having images matching the prescribed exam formats in Table 7-2. However, as previously discussed, the D1, D2, D3 fields in the TMS records used to store ovary dimensions often did not contain data in the correct order. Therefore it was not possible to validate the software by comparing directly the D1, D2, D3 dimensions obtained from the TMS with the D1, D2, D3 dimensions obtained from the images. Instead validation was achieved by comparing the ovary volumes for these exams using the two sources of dimension data. This avoided the problems of dimension ordering as volume could be calculated using the standard prolate ellipsoid formula ($\text{Volume} = D1 * D2 * D3 * 0.523$).

It was expected that both sources of dimension data should give the same ovary volume within a certain tolerance. This was because all the exams had been identified by the software as having images with calliper dimensions matching the dimensions recorded in the TMS (tolerance $\pm 1\text{mm}$). Accordingly, any significant difference in ovary volumes for a given exam would signal either problems with their images, a defect in the code used to process the images, or a problem in the subsequent processing of the data generated by the software.

The statistical software R (v3.2.0) was used to analyze the differences in ovarian volume between the two measurement sources. Correlation between the measurement sources for log ovary volumes in respect of LO and RO was determined using Pearson correlation coefficients and Scatter Plots. Agreement

between the measurement sources for ovary volumes in respect of LO and RO was determined using Bland-Altman analysis.

Calculation of Quality Control Metrics

QC metrics were calculated for each exam as defined in Table 7-3:

- **D1L** was set true if the image value of $D1 > D2$ and $D1 > D3$ for LO and RO.
- **Dimensions Assigned Correctly (DAC)** was set true if the difference between the D1, D2, D3 values obtained from the image and those recorded by the sonographer was ≤ 1 mm for LO and RO and the binary value ‘D1L’ was set also true for the exam.
- **D2O** was set true if the angle between the D1 and D2 calliper lines in the LS view of LO and RO was within the range ≥ 70 and ≤ 110 degrees.
- **D23** was set true if the ratio between D2 and D3 for LO and RO was in the range >0.5 and <1.5 .
- **Exam Result** was set true if the values of DAC, D2O, D1L and D23 for the exam were all set true.

Metric	Definition	Comment
D1L	Length of $D1 \geq D2$ and $D1 \geq D3$	$D3 > D1$ means D1 and D2 measured in TS which does not comply with UKCTOCS instructions.
DAC	D1L is true and dimensions of D1, D2, D3 recorded by sonographer match software measured values	Indicates that the ovary dimensions have been correctly assigned by the sonographer to values for D1,D2,D3.
D23	Ratio of D2 and D3 calliper line length, given D1L is true	Ovaries conform to a shape defined by D2, D3 dimensions. Error if outside range (0.5,1.5)
D2O	Angle between D1 and D1 calliper lines, given D1L is true	Error in angle results in error in volume calculation Error if outside range ($70^\circ, 110^\circ$)

Table 7-3: Novel quality control metrics for TVS exams of post-menopausal women were defined using the properties of the calliper lines found in the images.

Results

One thousand exams were randomly selected from the 68,931 TVS annual exams performed by UKCTOCS sonographers after 1st January 2008 that had their associated images stored in the URA and were also recorded as having both ovaries

a) visualized and b) classified as normal morphology as defined in Appendix B. These 1,000 exams had a total of 4,654 associated images (mean 4.65 images per exam, range 1-15). They were performed by a group of 96 sonographers of which 95 had performed more than one hundred TVS exams for UKCTOCS. Values recorded by the Sonographer for the left and right ovary dimensions (D1, D2, D3), as well as endometrium thickness, were imported into the image workflow IDF file from the TMS. The software reported that the TMS data was incomplete in the case of 39 exams (3.9% of total). Images associated with the remaining 961 exams were processed.

Identification of Images used to Measure Ovaries

The 4,654 images in the dataset were processed by workflow in 11 minutes 35 seconds. The exact images used to make all the measurements for an exam were identified in 55.1% (551) of the exams processed, so allowing recovery of this important information which had not been previously recorded by UKCTOCS. In the remaining 449 exams the exact image used to make each measurement could not be identified; 16% (160/1000) of exams did not record images of LO, RO and Endometrium, 16.4% (164/1000) had dimensions which did not allow the exact image to be identified as the values were too similar (unresolved), 8.6% (86/1000) had more than one image with the same calliper dimensions (duplicates). The results of processing the images associated with all 1,000 exams in the study are shown in Figure 7-4. It was possible to perform quality control for a significant collection of UKCTOCS TVS exams by analyzing the dimensions as described in Table 7-3.

Repeatability of Results

In order to confirm that the workflow software produced consistent results the images in the study dataset were processed twice with the same input parameters and processing steps. The output data for the D1, D2, D3 dimensions and D1-D2 angles were identical in each run. Therefore, the software was judged to exhibit 100% repeatability.

Manual Review

Thirty exams were randomly selected from the 551 exams in which the software had identified the exact images used by the sonographer to measure the ovaries. These exams were manually reviewed and it was confirmed the software had 1) correctly classified 100% of the exams in terms the image and exam formats as specified in Tables 7-1 and 7-2 respectively; 2) correctly ordered and measured the D1, D2, D3 dimensions from the calliper marks found in the images for 100% of the exams; 3) correctly measured the angle between the D1 and D2 lines with an error of less than ten degrees for 100% of the exams with differences between the manual and software measurements having mean of 2.47 degrees and a range of ± 6 degrees (SD=1.85). The above results suggest that the software was able to determine the D1L, DAC, D23 and D2O QC metrics with 100% accuracy.

Thirty exams were randomly selected from the 449 exams in which the software had not identified the exact images used by the sonographer to measure the ovaries. This second review confirmed that the software had classified with 100% accuracy the exams according to the image formats they contained (see Table 7-1). Therefore, the software was proven to correctly classify images for a random selection of exams in both subsets of the study dataset.

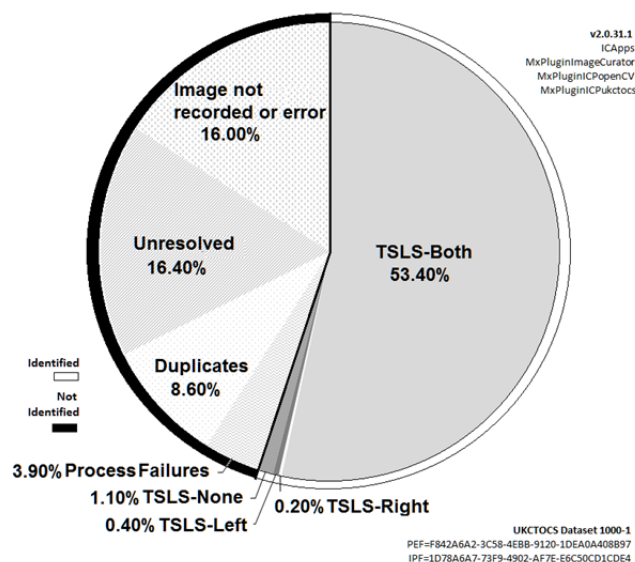


Figure 7-4: The success of identifying the exact images used to measure the ovary and endometrium from 1,000 TVS exams is reflected by the segments of the pie-chart contained within the white part of the outer circle described in terms of the examination formats; TSLs-Both, TSLs-Right, TSLs-Left, TSLs-None (Table 7-2).

Agreement of Ovary Volumes: Image vs TMS values

The D1, D2, D3 dimensions were converted into volumes to resolve the data ordering problem and so allow assessment of agreement between the data values recorded in the TMS for the exam and the data values obtained by the software from its associated images.

The calculation of Pearson correlation coefficients showed high correlation between log ovary volumes calculated from dimensions obtained from the software tool and sonographer recorded measurement sources; LO R=0.9893, RO R=0.9872. This high correlation between the two sources of values is also reflected in the scatter plots shown in Figure 7-5 (volumes displayed using linear scale).

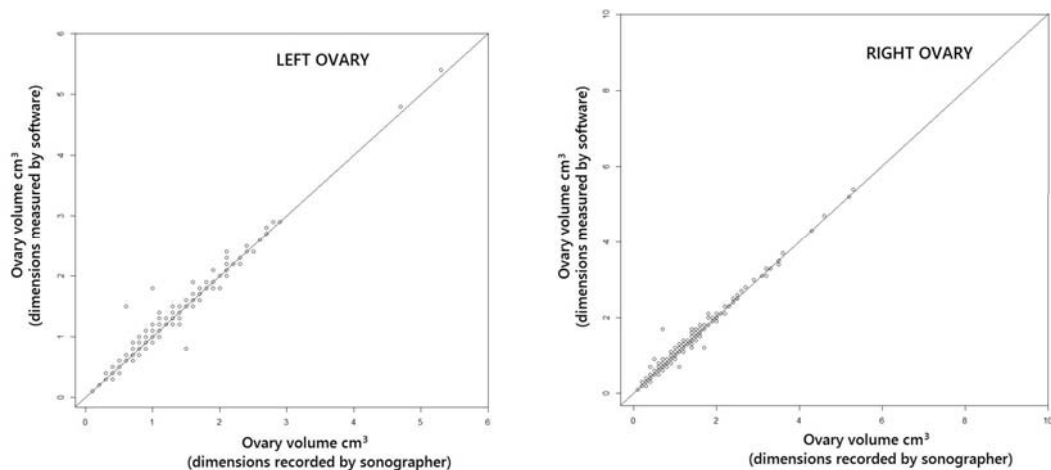


Figure 7-5: Scatter chart showing correlation of ovary volume calculated from sonographer recorded and software measured dimensions. The lack of significant outliers suggests that the software tool is operating correctly.

Bland-Altman analysis revealed good agreement between the ovarian volumes obtained from the software tool and sonographer recorded measurements; mean difference between the tool and sonographer values was -0.019 (LO) and -0.022 (RO) as shown in Figure 7-6.

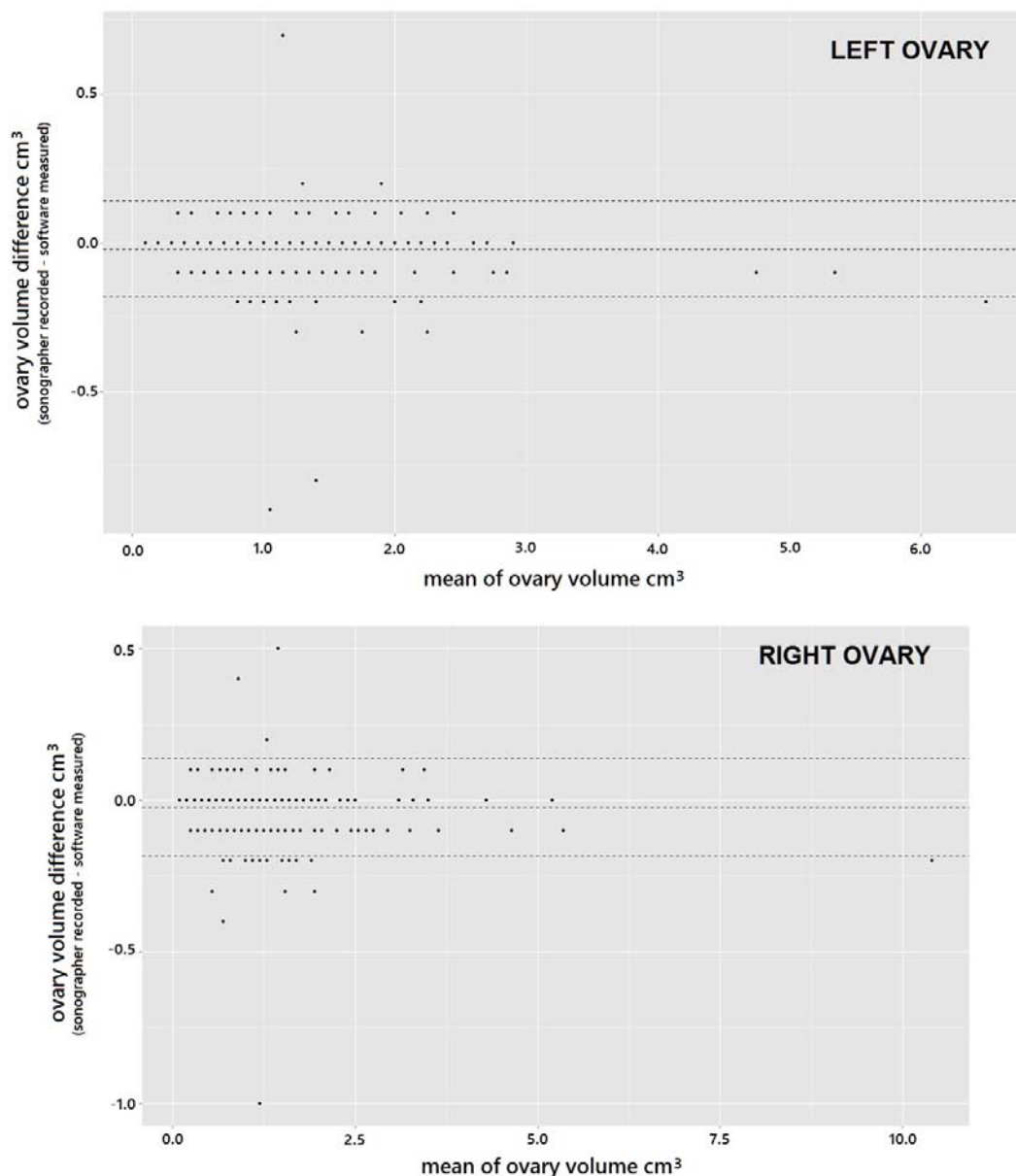


Figure 7-6: Bland-Altman plot shows agreement of sonographer recorded vs. software measured ovary volumes. The lack of significant outliers suggests that the software tool is operating correctly.

Calculation of Quality Control Metrics

The data from the 551 exams for which the workflow software was able to identify the exact images used by the sonographer to measure the ovaries was used to generate the QC metrics as defined in Table 7-3. The analysis set the D1L metric as true in 59% of exams, showing for these exams that the D1 dimension was equal or larger than the other two dimensions so confirming the sonographer had measured D1,D2 in the longitudinal section of the ovary as instructed. The DAC metric was set

as true for 49% of exams, confirming both D1L set as true and the recording of the dimensions in the correct order. The D23 metric was set as true for 90% of exams with D1L true, confirming the ratio of D2 and D3 was in the range typical for an ovary (0.5 to 1.5). Finally, the D2O metric was set true for 88% of exams with D1L true confirming the sonographer had placed D1 at right angles to D2 within ± 20 degrees for these exams (mean 94.4 degrees, range 55 to 132 degrees, standard deviation 8.75).

Each of these metrics corresponds to a different aspect of performing a good quality TVS exam, so they may be applied cumulatively to give an overall measure of quality. In the case of the 551 exams for which the metrics were calculated in this study, the percentage passing all four metrics D1L, DAC, D23 and D2O was 40%; see Table 7-4.

QC Metric	Description	Passing QC
DAC	D1L is true and dimensions of D1, D2, D3 recorded by sonographer match software measured values	49%
D2O	Angle between D1 and D1 calliper lines ≥ 70 and ≤ 110 degrees	88%
D1L	Length of D1 \geq D2 and D1 \geq D3	59%
D23	Ratio of D2 and D3 calliper line length > 0.5 and < 1.5	90%
Exam Result	Exams passing DAC, D2O, D1L and D23	40%

Table 7-4: Results of calculating the DAC, D2O, D1L and DCR metrics using data from the exact images used to measure both ovaries in the case of the 534 TVS exams (53.4%) identified as having the format TSLS-Both.

Discussion

A software tool has been developed which allows the collection of novel QC metrics by analyzing the images used by the sonographer to measure ovaries during TVS scanning. These metrics are objective, free from inter/intra observer variability and measured independently of the sonographer at low cost. The tool is highly accurate and produces repeatable results. It has also revealed that D1-D2-D3 ordering in TMS is not reliable, so values obtained by the tool should be used when ordering is important, for example when calculating some QC metrics for UKCTOCS TVS exams.

It is recommended that future trials involving TVS employ a similar software tool to gather QC metrics from the exam images and use them to drive improvements in scanning as part of a quality improvement programme. Care should also be taken to identify the exact images used by the sonographer to measure to ovaries so that such metrics can be determined by the software tool for 100% of exams. Unfortunately, in the case of TVS exams performed by UKCTOCS the sonographer was not asked to identify such images, so identification could only be achieved for just over half of the exams in the study dataset.

Identification of the Exact Images Used for Measurement

The recovery of the information needed to identify the exact image used by the sonographer to measure the ovaries was a significant problem for the software to address due to the large number of images that needed to be processed as well as the wide variation in their formats and calliper types. In addition, the bug found in the measurements made by the ultrasound machine meant the values did not exactly correspond to the length of the associated calliper line in the image. Consequently, the number of the exams reported as unresolved (16.4%) or duplicates (8.6%) was larger than would have been the case had it been possible to match the measurements exactly rather than within a set tolerance ($\pm 1.0\text{mm}$). The alternative approach of using optical character recognition to read the values displayed at the bottom of the image may have resulted in more images being identified, but would have presented additional errors due to the variation in font, location and format of the numbers.

Capability of the Workflow Software as a Measurement Device

The validity of the software is demonstrated by 100% repeatability and 100% accuracy of the data it generated when compared with values from manual review. Furthermore, the low mean difference between software tool and sonographer recorded measurement sources indicates good agreement between them; LO mean difference = -0.0190 , RO mean difference = -0.0229 . The Bland-Altman (BA) plot (figure 7-6) shows the mean difference between ovary volumes calculated from sonographer recorded and software measured dimensions for LO as -0.019 (95% CI: $-0.181, 0.143$) and for RO as -0.022 (95% CI: $-0.184, 0.138$). However, there were a few outliers with a larger volume difference than would be expected for dimension

match tolerance ± 1 mm. This illustrates the value of the BA plot in identifying such potential problems. The cause of these exceptions will be investigated before the next release of the software, but their low incidence suggests they may result from the processing of images with unusual properties.

It appears that there is a slight positive systematic bias for software values which may be due to a small defect that has been observed in the way the ultrasound machine calculates the length of calliper lines, as described previously. However, the distribution of differences seems to keep around zero even as the mean of the two method values increases, so there does not seem to be a proportional bias. Therefore it can be concluded that the software is a highly capable measurement system.

Limitation of Study

A weakness of this study is that it only considers the capability of the software tool for the 55.1% of exams with images that can be precisely identified as having been used by the sonographer to measure the ovaries. However, a later study (Chapter 9) found good agreement between the visualization rates obtained by expert review in both subsets, which suggests that no confounders or bias exist between them in respect of image quality.

Quality Control metrics Applied to UKCTOCS TVS Scanning

The need to calculate metrics using information obtained from images rather than recorded in TMS is highlighted by the low percentage (49%) of exams with Dimensions Assigned Correctly (DAC) by the sonographer. This suggests values for D23 and D1L generated from UKCTOCS TMS data might not be reliable as they depend upon the correct identification of D1,D2,D3 though correction could be applied in other ways. In addition, exams with the metrics D1L and D2O set true may be used as criteria for future studies as errors in the placement of the calliper marks will impact the accuracy of D1, D2, D3 values.

Summary

This chapter has outlined some of the challenges associated with processing images from TVS examinations performed by UKCTOCS. It has also described how the workflow software described in Chapter 6 was adapted to process 4,654 images from 1,000 TVS exams in less than twelve minutes with high accuracy. This resulted in identification of the exact image used to measure the ovary in the case of 551 exams (55.1%). It also recovered the ordering of the D1, D2, D3 ovary dimensions and D1-D2 angle from these exams so permitting the calculation of QC metrics that are both objective and independent. Therefore it can be concluded that the workflow software, together with the custom library developed for this study, form an effective tool for automating the analysis of static images generated by TVS scanning and generate useful QC metrics at low cost.

A further study in Chapter 9 uses the same tool and image dataset to explore the relationship between ovary dimension values and visualization rates from reviews of the same exam images by experts in gynaecological scanning, so providing further evidence to answer the research question posed in Chapter 1.

Chapter 8: Audit of TVS Examinations by a Team of Experts

Introduction

A key metric for the quality control (QC) of TVS scanning is visualization rate (VR), as discussed in previous chapters. In this chapter the definition of VR is further refined as VR-Both; the proportion of TVS exams where both ovaries are seen, or exams where the remaining ovary is seen in the case of women who have undergone a previous unilateral oophorectomy. This reflects the significance of the sonographer inspecting both ovaries for morphological abnormalities¹³ when screening for ovarian cancer, though it must be accepted that the ovaries of postmenopausal women generally become smaller with age and more difficult to visualize as reported by Sharma et al³⁰.

The ability of sonographers to visualise small ovaries / adnexal masses has particular importance when TVS examination is performed as a screen for ovarian cancer due to the evidence that suggests high grade serous ovarian cancers remain less than 1 cm in diameter for up to 4 years before they present clinically, as reported by Brown and Palmer²⁵. The model they developed suggests an annual screen would need to detect adnexal lesions as small as 1.3 cm in diameter in order to pick up 50% of high grade serous ovarian cancer before they spread and become stage III; see discussion in Chapter 1. Further evidence in this regard is provided by data from ovarian cancer screening trials in respect of scans finding only normal ovarian morphology in the year prior to diagnosis of Stage III Type II invasive epithelial ovarian cancers, as reported by Sharma et al²⁰⁶ as well as Partridge et al^{207,208}. Therefore a prerequisite for detecting the type of small adnexal lesions suggested by Brown and Palmer is the visualization of both ovaries during TVS scanning. Consequently the metric VR-Both has relevance as it reflects sonographer ability in this regard.

UKCTOCS formed the Ultrasound Subcommittee (USC) in 2006 and also implemented an accreditation programme, which amongst other objectives required each sonographer to achieve VR-Both greater than 60% over a 3 month period. The

USC reviewed the performance of UKCTOCS sonographers over the next two years and observed a small number had self-reported VR-Both of over 90%. Therefore in 2009 it was decided to audit the performance of these high scoring sonographers to confirm independently that it is possible to achieve ovary visualization rates of greater than 90% in postmenopausal women. This chapter describes the audit and its outcome. The findings are being prepared for submission for publication in *Ultrasound in Obstetrics & Gynecology*²⁰⁹. The co-authors collaborated in the design of the study and the generation of the data as well as its interpretation. All authors contributed to the editing of the paper, parts of which are reproduced in this chapter.

Methods

Interrogation of the Trial Management System (TMS) data allows calculation of a sonographer's visualization rate from binary fields in the Scanning Report Form relating to whether the ovaries were seen or not, as shown in Appendix C. Such data is supported by the presence of the ovary dimensions (D1, D2, D3) as reported by the same form, which signifies not only that the sonographer had visualised the ovary, but also that she had captured the static image(s) necessary to measure it. Therefore an audit of the sonographer's self-assessed VR-Both can be performed by reviewing these images to provide independent confirmation that the feature measured was an ovary.

An audit was performed between July and September 2009. It involved seven sonographers who had recorded the highest results for VR-Both between January 2008 to January 2009 and who had also performed more than 100 TVS exams during this period. The results of the audit were reported to the UKCTOCS Ultrasound Subcommittee (USC) in November 2009. The audit dataset of 357 exams was created by randomly selecting 51 annual exams performed by each of the seven sonographers during the study period which had the corresponding images archived in the URA. All seven sonographers used the same ultrasound machine; a Medison Accuvix XQ as described in Chapter 2. In addition all the sonographers were accredited by UKCTOCS as described in Chapter 3.

Exams were included in the dataset if both ovaries were reported as visualised and they were classified as having normal morphology according to the UKCTOCS protocol as defined in Appendix B. Exams were excluded if classed by the protocol as unsatisfactory, abnormal, or if the exam had one or more ovaries categorized as 'good view'; classifications that are also defined in Appendix B. The classification of exams was performed by the TMS using data entered from the sonographer's Scanning Report form. This classification algorithm was derived from the information reproduced in Appendix B and is also described by the flow diagram contained in Appendix D.

The nine auditors selected to review the images in the audit dataset were all highly experienced in gynaecological scanning, particularly TVS. They were comprised of

four consultant gynaecologists, two gynaecological radiologists specialising in gynaecology and three sonographers employed at superintendent grade in the NHS. All were members of UKCTOCS Ultrasound Subcommittee (USC). The auditors were split into three groups; A, B, C. Two groups had a consultant gynaecologist, a radiologist and a sonographer. One group had two consultant gynaecologists and a sonographer. Therefore all groups contained auditors with different levels of training and experience. However, it was noted that two of the auditors in group B regularly worked together to perform reviews of TVS images. The dataset was partitioned by sonographer and then each partition was randomly assigned to the three groups of auditors (A,B,C); see Figure 8-1.

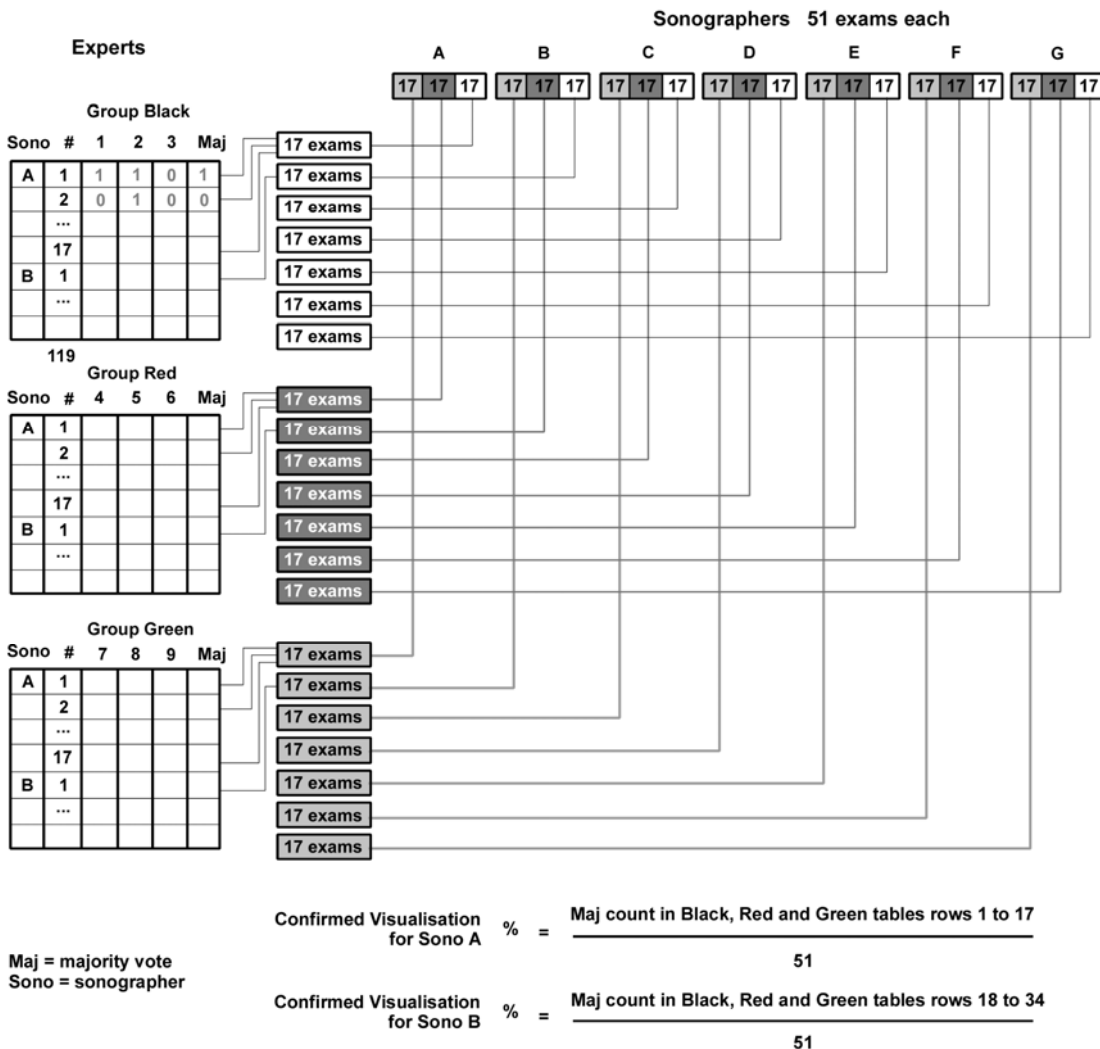


Figure 8-1: Partitioning of sonographers' TVS exams to the reviewers

Each auditor worked independently to review all the images from each of the exams assigned to their group in order to reach a conclusion about whether both ovaries had been measured by the sonographer. This involved making a decision about whether or not the object in the image marked by callipers was an ovary. The decision was considered a matter of individual subjective judgement, particularly as the auditors had not previously agreed criteria such as acceptable ranges of size and shape for normal ovaries. Conclusions about a particular exam were reached both by majority verdict (two of the three auditors in a group agreeing) and by unanimous verdict (all three auditors agreeing). In this way, each group reviewed one third of the exams performed by each of the seven sonographers, so conclusions about each would reflect the consensus of all nine auditors.

The auditors were aware that the exams had been randomly selected from all those performed by each sonographer over a twelve month period that had been recorded as both ovaries seen and having normal morphology. Therefore they knew the exams were representative of the work that had been used to calculate the sonographer's VR-Both. The primary aim of the audit was to confirm the VR-Both for each of the seven sonographers. The auditors were also required to identify the exact images used to measure the left and right ovary (LO, RO) from all of the images captured during the exam. In order to facilitate this process a software application called osImageManager was adapted for use by the auditors, as described in Chapter 5.

The osImageManager adaptations had the aim of a) displaying the images associated with an examination under review in a user-friendly GUI and b) allowing review results to be captured reliably in a way that facilitated their subsequent analysis. A separate review database was created from data in the URA for each of the groups (A, B, C) with each containing the images from the 119 exams randomly allocated to the group. The software displayed the exams and associated images from the given database (A, B or C) and provided forms for recording of the review assessment. In this way the auditors could efficiently work through the list of exams, recording data about the images that had been successfully reviewed.

The audit was prepared by installing osImageManager and the review database (A, B, or C) on a number of laptop PCs, though external monitors (22") were used to

display the images during the review. The auditors were each trained in the use of the system by the author before starting their review.

Statistical analysis

Primary comparison was based on the agreement between the auditors assessment and sonographer's report in terms of both ovaries being seen for the examination; *VR-Both Confirmed*. Mean, range and standard deviation values were also calculated from this data so auditors producing outlying results might be identified and their contribution to the study removed from the audit database if necessary.

The modified audit database was queried to produce graphs allowing comparison of sonographers in terms of the percentage of exams that two or three of the three auditors in a group had agreed with the sonographer's assessment that both ovaries were seen, on the basis of their review of the static images used to measure the ovary; majority vote for *VR-Both Confirmed*. Values for mean, range and standard deviation of the majority votes were also calculated for the sonographers combined *VR-Both Confirmed*. The analysis was repeated using the criterion of all three auditors agreeing with the sonographer's report; unanimous vote for *VR-Both Confirmed*. The baseline characteristics of the women (trial centre code, age, years since last period, body mass index (BMI), hysterectomy status, OCP and HRT use) were compared to detect any bias in the results of the audit.

Information from the UKCTOCS sonographer accreditation records was used to calculate the mean, range and standard deviation of their collective experience as well as the number of different trial centres in which they were based. Their level of training and qualifications was also compared. The time each auditor took to review the images was recorded and analyzed to calculate the number of man-hours needed to complete the audit.

Kappa Statistic

The Cohen Kappa²¹⁰ statistic is often used in studies that involve subjective interpretation by observers in order to take account of agreements occurring simply by chance. For example, if two similarly skilled and experienced technicians working in the same lab are making a decision about the presence or absence of a particular smell in a batch of test tubes, then close agreement might be expected in

terms of which tubes have the smell and which do not. If the work was repeated when the technicians had heavy colds causing them to guess the result, then agreement would still happen in some instances by chance, but the agreement would not be as close. In such circumstances the Kappa statistic would be useful in terms of indicating a problem with the data. Viera and Garret²¹¹ explain the Kappa statistic as a measure of precision of the agreement between two observers. In this respect they differentiate between accuracy and precision by making an analogy to target shooting like that shown in Figure 8-2.

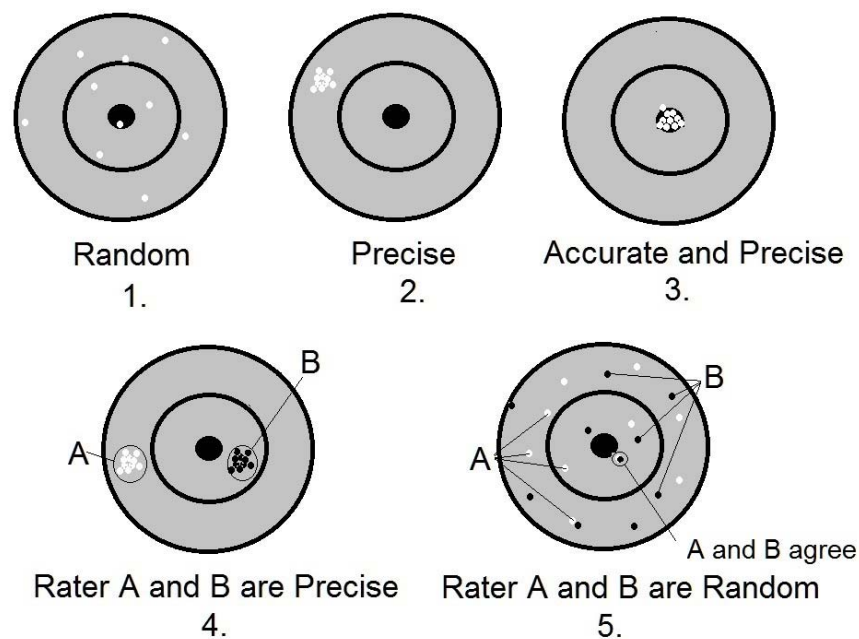


Figure 8-2: Accuracy and Precision of marksmen shooting at a target. The first target has a random distribution of bullet holes suggesting the marksman is not precise, though by chance he hits the bulls-eye once. The second target has a close grouping of bullet holes suggesting the marksman is precise, but not accurate as he doesn't hit the bulls-eye once. The third target has a close grouping of holes in the bulls-eye so the marksman is both accurate and precise. The fourth and fifth targets results from two marksmen shooting at the same target. In the case of the fourth target, they are both precise but not accurate, but in the fifth target they are randomly hitting the target. The Kappa statistic would be useful when attempting to provide a measure of agreement between marksmen in the case of the fourth and fifth target.

The Kappa statistic was calculated for pairs of experts in the same group so three sets of values were produced for each group; nine values in total. This would measure agreements occurring between experts just by chance. Preparation for the calculation involved arranging the results produced by each pair of experts into 2x2

tables containing totals for the first expert agreeing with the second expert in terms of exams with both ovaries visualised or not visualised as well as disagreements for the same categories. The calculation itself was performed by Matthew Burnell (co-author of the paper resulting from the study) using Stata statistical analysis software²¹².

The interpretation of Kappa value was provided by Landis and Koch²¹⁰ in terms of agreement being: a) less than chance (<0), b) slight (0.01-0.20), c) fair (0.21-0.40), d) moderate (0.41-0.60), e) substantial (0.61-0.80) f) almost perfect (0.81-0.99).

Results

The audit dataset was formed by images from 357 annual TVS exams on 349 women performed by seven UKCTOCS sonographers during the study period; 1st January 2008 and 31st December 2008. These sonographers were selected on the basis of having the highest self-assessed VR amongst all the 96 sonographers employed by UKCTOCS during this period and all had VR-Both that was greater than or equal to 89%.

The dataset contained 51 exams performed by each of the seven sonographers according to the UKCTOCS protocol and were selected at random from all those they had individually undertaken during the study period and that had been archived in the URA. Exams were included if classified as normal according to the UKCTOCS protocol and both ovaries were reported as visualised. Exams were excluded if classed by the protocol as unsatisfactory, abnormal, or if the exam had one or more ovaries categorized as 'good view'. The exams in the dataset had a mean of 5.4 images per exam (range 1-30) and the same ultrasound machine was used in all cases.

Each group of three auditors were expected to review 119 of the total 357 exams and each auditor was expected to review 17 exams performed by each of the 7 sonographers.

In 2010 it was discovered that auditor 9 was also sonographer C, whose work was being audited. All the experts were members of the Ultrasound Subcommittee and

had been involved in the planning of the audit. Therefore they were aware of the sonographer selection criteria (VR greater than or equal to 89% in 2008). All sonographers were also aware of their own VR results for 2008. It is presumed that this conflict of interest was not identified due to an innocent oversight. However, to avoid the possibility of observer bias the results from auditor 9 were removed from the audit database. Consequently all reviews in the same group (C) also had to be removed as it would no longer be possible to calculate a majority vote and the criterion for unanimous vote would be different in this group (2 of 2, rather than 3 of 3).

The decision to remove the results of group C from the audit database was validated by a comparison made of the 9 auditors' work in terms of the percentage of exams they had agreed with the sonographer's assessment that both ovaries were seen; see Figure 8-3. These figures for 'VR-Both Confirmed' have mean=69.86%, range 46.61% to 96.61%, SD=16.30. It was noted that auditor 9 had confirmed ovary visualization in 96.61% of exams he/she reviewed and was therefore considered an outlier in respect of the results of the other auditors.

In addition to the removal of all results from one group, the auditors in the remaining groups failed to review all images in the case of 15 exams. Therefore the total number of exams reviewed for all 7 sonographers by the six auditors in the two remaining groups was 223. Therefore the mean number of exams reviewed for each sonographer was 31.86 (range 22 to 42, SD=6.77).

The results of the auditors' reviews were analyzed by majority vote and unanimous vote for VR-Both Confirmed. When majority voting was applied either two or three of the auditors in each group had agreed the outcome of their review of the images for a particular exam. Aggregating such outcomes from both groups of auditors for all 223 exams performed by all seven sonographers gave a mean 'VR-Both Confirmed' of 70.05% (range 53.57% to 91.89%, SD=14.73); see Figure 8-4. Repeating the analysis using the criterion of unanimous vote (all three experts in the group agree) gave a mean 'VR-Both Confirmed' of 47.34% (range 24.14% to 70.27%, SD=15.79).

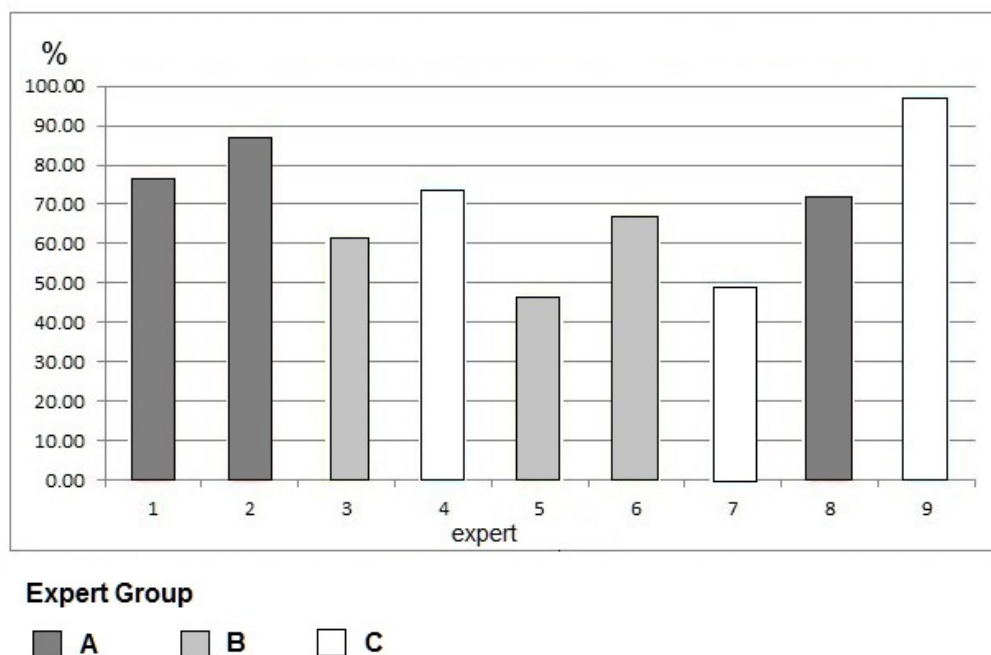


Figure 8-3: Variation in VR-Both Confirmed by individual auditors shows auditor 9 agreed with the sonographer’s judgement in almost all cases where as auditor 5 agreed in less than half of the exams he reviewed. The variation between the confirmation rates of the various auditors may reflect their different training and experience.

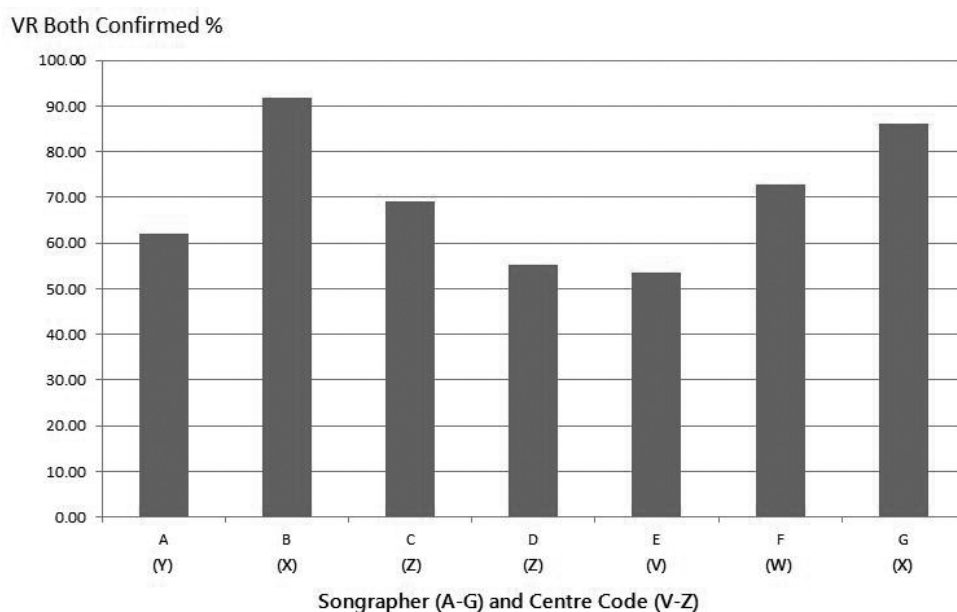


Figure 8-4: Confirmation of sonographer VR by the team of 6 auditors in two groups such that each group reviewed half of each sonographers’ exams and each exam was reviewed by the three auditors in the group with a decision reached by majority decision. Group C auditors were excluded. It should be noted that all six auditors seem to agree with the self-assessment of VR made by sonographer B (92%) on the basis of images reviewed from 32 exams.

Further analysis of the individual sonographer data revealed the existence of two clusters for 'VR-Both Confirmed' (majority vote) such that one group of 2 sonographers (B and G in Figure 8-4) had a mean of 89% (range 86%-92%) and another group of 5 sonographers had a mean of 62% (range 54%-73%).

Volunteer Variation

Sharma et al³⁰ identified a number of factors that affected visualization of postmenopausal ovaries. Therefore an analysis was performed on the 349 women involved in the study to identify the proportions of these factors in the women scanned at each trial centre to identify whether any sonographer might have had a disadvantage in terms of the type of women they were scanning; see Figures 8-5 and 8-6.

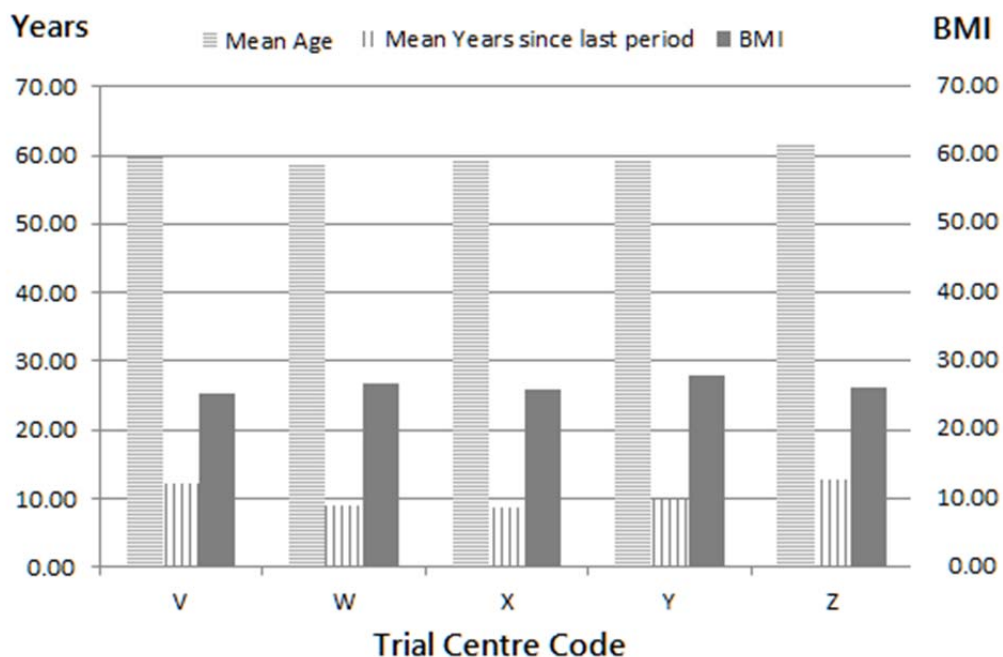


Figure 8-5: At randomisation the women in the study had a mean age of 60.0 years (range 50–73, SD 5.8), a mean time since last period of 10.7 years (range 0.0–42.7, SD 8.0) and a mean body mass index (BMI) of 26.3 (range 17.5– 68.5, SD 4.7). Analysis by trial centre suggests that the sonographers examined similar numbers of women for whom ovary visualization might be considered difficult due to their age, time since last period or BMI.

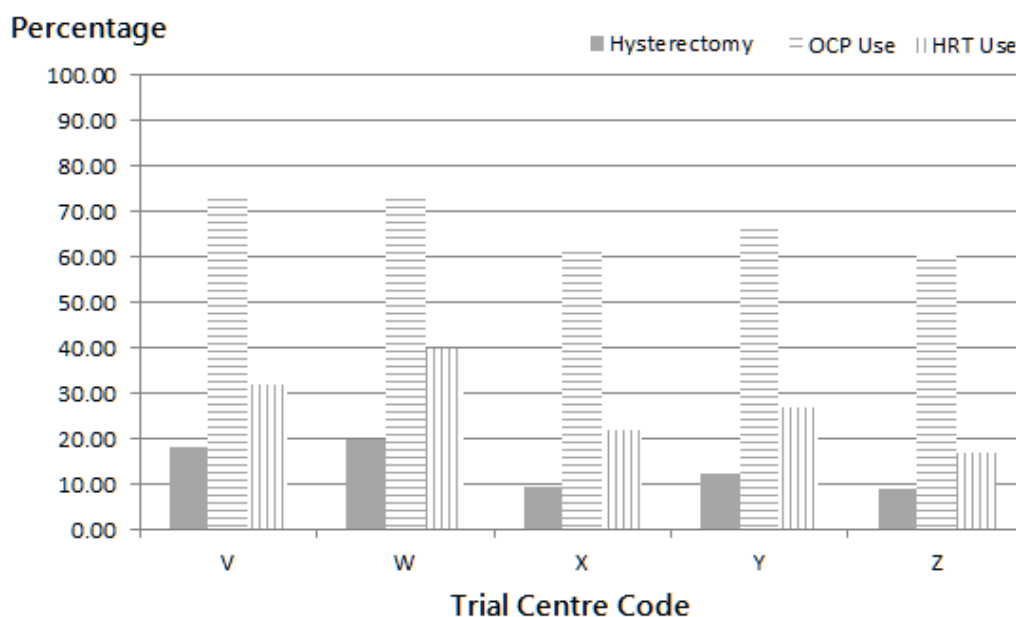


Figure 8-6: At randomisation 24.9% of the women in the study were using hormone replacement therapy (HRT), 64.8% were using oral contraceptive (OCP) and 12.3% had had a hysterectomy. Analysis by trial centre suggests that the sonographers examined similar numbers of women for whom ovary visualization might be considered difficult due to hysterectomy status or use of OCP or HRT.

The sonographers had a mean experience of 14.5 years (range 7 – 23, SD 7). They operated in 5 different trial centres with 2 sonographers working in centre X and 2 sonographers working in centre Z. All sonographers had the same level of qualification, had completed similar training before starting work on the trial and were accredited by UKCTOCS during 2008.

Observer variation between auditors

A number of statistics were calculated to assess observer variation between the various pairs of auditors in each group. The results are summarised in Table 8-1.

Resources required by the audit

The nine auditors performed the image review at locations in Manchester, Derby and London. Collectively they spent approximately 150 hours performing three reviews of images from 51 exams by each of the seven UKCTOCS sonographers.

Reviewer Pair	Expected Agreement	Agreement	Kappa	Std. Err.	Z	Prob>Z
A						
1, 2	77.31%	69.80%	0.2488	0.0855	2.91	0.0018
1, 8	73.33%	61.56%	0.3064	0.0905	3.38	0.0004
2, 8	77.12%	65.80%	0.3310	0.0812	4.07	0.0000
B						
3, 5	70.59%	49.33%	0.4195	0.0880	4.77	0.0000
3, 6	77.31%	53.91%	0.5077	0.0909	5.58	0.0000
5, 6	73.11%	48.99%	0.4729	0.0842	5.62	0.0000
C						
4, 7	58.47%	50.00%	0.1695	0.0819	2.07	0.0192
4, 9	74.58%	71.33%	0.1132	0.0517	2.19	0.0142
7, 9	51.26%	49.61%	0.0328	0.0328	1.00	0.1586

Table 8-1: Observer variation between pairs of expert reviewers in each of the three groups A, B, C. It can be seen the best agreement is between auditors 3 and 6 in group B (0.5077 or moderate) and the worst agreement is between auditors 7 and 9 in group C (0.0328 or slight).

Discussion

The audit concluded that the self-assessment of ovary visualization by these seven sonographers was unreliable. Almost a third of the audited exams were judged unsatisfactory because the auditors considered structures other than left and right ovary had been mistakenly measured, suggesting that both ovaries had not been visualised. However, individual sonographer performance varies significantly in this regard, with one group of two sonographers (B and G in Figure 8-4) correctly visualising both ovaries in almost all their reviewed exams. These results suggest that it is indeed possible for sonographers to correctly visualise both ovaries when scanning menopausal women, regardless of factors considered to make ovary visualization challenging such as old age and high BMI³⁰.

Factors causing variation in VR-Both amongst sonographers

The data was not considered sufficient to determine whether the variation in VR-Both reported by individual sonographers was associated with any clustering of factors known to influence ovary visualization. However, a similar distribution of such factors was identified across all centres in which audited sonographers worked and is consistent with results reported elsewhere for UKCTOCS as a whole².

Furthermore the standardisation of the ultrasound machine, training and scanning

process (which had been completed by January 2008) reduces the likelihood of variation in VR-Both being associated with factors related to the scanning centre. Therefore it can be concluded that the variation in VR-Both achieved by these individual sonographers results from differences in ability. This supposition is supported by findings reported by Sharma et al², which conclude that major contributors to VR-Both rates are individual sonographer skill, attention to detail and experience.

Observer variation between the expert reviewers

The initial impression created from inspection of the Kappa values is that the significant variation in inter-observer agreement between the auditors in each group casts doubt on the validity of the audit. At best the auditors in group B had moderate agreement and at worst the auditors in group A had fair agreement. This suggests that in a significant number of cases they could not agree whether or not an exam had visualised both ovaries after reviewing its images. However, there are some reasons for caution before reaching such a conclusion.

In the case of the audit of UKCTOCS sonographers, the Kappa statistic is measuring the precision of two auditors' opinion about the presence or absence of both ovaries in images from a TVS exam. It does not measure the accuracy of this opinion. Furthermore it cannot determine if both experts are precise but differ in accuracy; the case of target 4 in Figure 8-2. Kappa is limited to identifying agreement that arises by chance such as the case of two poor marksmen putting a bullet through the same hole in the case of target 5 in Figure 8-2. In order for Kappa to have any significance in respect of this audit it is necessary to assume that some of the auditors were making their assessment entirely by chance. This seems unlikely. It is more likely they were making their assessments using different criteria (subconsciously or otherwise) so their precision was similar, but frequently they did not agree; the case of target 4 in Figure 8-2. This is a reasonable hypothesis given that only auditors 5 and 6 had worked together assessing TVS images and they achieved the second highest Kappa score in terms of agreement at 0.473 (moderate).

The need for two auditors to have an equal probability of assessing that both ovaries had been visualised during an exam is the assumption of marginal homogeneity

between raters as identified by Zwick²¹³. It is noted by Parpia et al²¹⁴ that this assumption is violated if there are differences in training and experience between raters, as was the case in for the auditors in this study. Indeed members of the auditor groups were selected on the basis of their training and experience with the aim of achieving diversity in each group ranging from a superintendent ultrasonographer to consultant gynaecologist. Whilst such differences may not necessarily reflect skill at the interpretation of TVS images, the possibility of some marginal heterogeneity between auditors existing in the case of the audit has to be acknowledged. Therefore caution is needed before challenging the results of the audit on the basis of there being only fair or moderate agreement between auditors as determined by the calculation of Kappa values. This is because the use of such values to evaluate inter-auditor agreement may not be appropriate for this type of study as suggested by Zwick²¹³, Parpia et al²¹⁴ and Viera and Garret²¹¹; see Limitations of study for a further discussion.

Results compared with other work

The results support the re-classification of more than a third of the TVS exams performed by the group of 5 sonographers which were audited as unsatisfactory (Figure 8-4, A, C, D, E, F). This compares with other work performed on the Quality Control (QC) of UKCTOCS TVS scanning with similar exam selection criteria which provided evidence for re-classifying as unsatisfactory approximately half of exams performed by 96 sonographers reporting VR-Both in the range 60% to more than 90%; see Chapter 9.

Planning and execution of future studies

The problem of intra-observer bias was not addressed in our study, though it would have been possible to account for such variation by randomly inserting repeats of the same exam in the review sequence and then measuring agreement. It is suggested that future VR-Both studies adopt this approach and also confirm that an auditor is able to demonstrate a low level of intra-observer variation before participating. Information about the identification of the sonographer responsible for creating the image was not given to the auditors during the image review to avoid bias arising from any personal association with the person whose work they were reviewing.

Future studies should also attempt to address any bias arising from group association between the auditors and the people they are assessing.

Limitations of the Study

The auditor groups were set up to provide judgement about ovary visualization from auditors with a range of training and experience. It seems likely that this lack of homogeneity amongst the auditors was more significant in terms of inter-observer agreement as measured by the Kappa scores given in Table 8-1 than the effect of random guesses being made. That is to say agreement between auditors was the result of them having different opinions about accuracy as a result of their different backgrounds rather than any lack of precision associated with poor skill. The analogy in terms of Figure 8-2 is two auditors not agreeing on the location of the bulls-eye so grouping shots at different points on the target (4) rather than shooting randomly and sometimes hitting the same point by chance (5). However, the Kappa statistic is only a measure of the latter and not the former. Therefore caution is needed in the interpretation of the Kappa statistic to quantify inter-observer agreement in this audit. Accordingly the low Kappa values found between pairs of auditors does not justify the findings of the audit being questioned on these grounds, though there is a strong suspicion that the differences in the accuracy of some auditors' results may have reduced or increased the mean values of confirmed VR for the sonographers. In this respect the study is limited.

Intra-observer variability was not addressed by the study so the capability of individual auditors to provide consist results for the same exams was not measured. Had this been done then some weighting of the auditors' results may have been performed so permitting compensation for differences in their accuracy for assessing ovary visualization. In this way a more reliable indication of inter-observer variability may have been obtained, though the challenge of identifying the most accurate auditors would remain a significant limitation of the study.

Future Work

The way the auditors reached their decisions about whether or not the sonographer had seen both ovaries during their TVS examinations for UKCTOCS is poorly understood. It is known that such judgements are subjective, but there is an

expectation that a group of auditors with similar levels of training, experience and skill will reach a consensus. Furthermore some of the criteria in respect of ranges of size and shape for normal ovaries have been codified so that the VR for a collection of examinations may be obtained by a computer program as described in Chapter 9. This suggests that at least part of the subjective decision may result from expectations about objective factors such as ovary size and shape. However, the auditors did not share quantitative information in this regard. In addition, other criteria about the object in the image measured as an ovary by the sonographer may have been influential, such as the texture of the object, its boundary and content. Therefore it can be concluded that most of the auditors used different criteria to reach their judgement based on their training and experience.

Future research should be undertaken to investigate the impact of auditors training, experience and skill on inter-observer agreement when reviewing images from TVS exams to decide whether or not the sonographer had measured an ovary. This would provide information that might allow the weighting of the Kappa statistic calculation for observer expertise and so give more reliable values for future studies similar to this audit. Future studies may also benefit from measuring intra-observer variability in respect of individual auditors. This would allow some measure of the precision of their assessments. However, in order to obtain a better idea of their accuracy it would be necessary to engage a large number of auditors able to produce precise assessments so that a consensus might be reached about which exams had both ovaries visualised and which exams did not.

Summary

The audit indicates assessment of normal ovary visualization by most of the sonographers is unreliable, as almost a third of exams were judged to be unsatisfactory because the expert reviewers were not convinced both ovaries had been visualised. However, the proportion of such exams may be greater or less than a third due to the observer variability between the auditors which could not be quantified by this study. Indeed the study design compounded the problem by mixing auditors with different training and experience within the same review group so making use of the Kappa statistic problematic. Irrespective of such inter-observer variability and its impact on the audit as a whole, there is evidence that a few sonographers may be able to visualise both ovaries in almost all cases when scanning a range of post-menopausal women. This is suggested by the 92% VR-Both Confirmed for sonographer G which indicates a high level of agreement between all auditors in respect of this particular sonographer.

Although there are other reports of reviews involving TVS images^{215,216} none have been conducted in a systematic way by a panel of expert reviewers for a group of sonographers in order to audit VR-Both. The value of this study is that has raised concerns about the reliability of this metric when produced from information about ovary visualisation provided by the examining sonographer. The study also identifies the difficulties of performing such a review and suggests ways in which future studies might improve their outcome. However, a key finding of the study is the suggestion that visualisation of both ovaries is possible for most post-menopausal women during TVS examination.

The ovary of a postmenopausal woman may sometimes not be seen during a transvaginal ultrasound (TVS) exam as ovaries typically shrink with age to the point that they can be very difficult to locate using ultrasound. It is for this reason that UKCTOCS allows an exam to be classified as having a normal outcome if both ovaries are not seen, but a good view has been achieved of the iliac vessels. However, where both ovaries are present, it would seem most exams have the potential to succeed in visualising them. This is important when TVS is used in screening for ovarian cancer, for otherwise abnormalities may not be detected.

Further research would be required to determine if the visualization performance of individual sonographers results from innate skill or is a product of better training, mentoring or experience in TVS scanning. Should better performance be explained by innate skill then recruitment policies may need to be adapted, otherwise a Quality Improvement (QI) programme might help more sonographers achieve high levels of ovary visualization. In the latter case in order to allow targeting of support effectively alternative approaches to the quality control of TVS are needed to replace the current practice of sonographers self-reporting VR-Both. Research to find alternative metrics that can be obtained at low cost and which give an independent and objective measurement of sonographer performance when TVS scanning is used as a screen for ovarian cancer is described in Chapter 9.

Chapter 9: Automated Collection of TVS Quality Control Metrics

Introduction

Visualization Rate (VR) has been previously discussed as the key quality control metric used in TVS scanning. A number of cancer screening trials involving TVS have measured VR; University of Kentucky ovarian cancer screening study²¹⁷, Prostate Lung Colorectal and Ovarian (PLCO) cancer screening trial¹²⁴, UKCTOCS². Various studies have also reported VR^{218,219,220,221}. In all cases VR was entirely self-reported by the examining sonographer, with the exception of the PLCO trial where visualization was confirmed by a different sonographer during the same participant visit for a small percentage of exams. The reliance on the self-reporting of VR reflects the significant difficulty of obtaining objective and independent Quality Control (QC) metrics for TVS at low cost.

The study reported in the previous chapter found that the assessment of ovary visualization by the examining sonographer was unreliable in five out of the seven sonographers whose work was audited by a team of experts. This highlighted the need for further research to find alternative QC metrics to give an independent and objective measurement of sonographer performance when TVS scanning is used as a screen for ovarian cancer.

This chapter reports on a study to investigate the production of a novel QC metric calculated from values obtained by software analysis of static images from TVS exams. The study also involved an expert review of the same collection of images to provide an independent measure of VR. The results achieved by software analysis were compared to visualization assessed by expert review in order to evaluate the potential of the metric for QC purposes. In this way it was demonstrated that a software tool might help in the collection of independent and objective QC metrics for TVS exams, so providing information to help answer the research question posed at the start of this thesis. The findings of the study have been submitted for publication in *Ultrasound in Obstetrics & Gynecology*²²². The co-authors

collaborated in the design of the study and the generation of the data as well as its interpretation. All authors contributed to the editing of the paper parts of which are reproduced in this chapter.

Background and Objectives

In the case of a static TVS image used to measure an ovary, the callipers indicate the boundary extents of the object considered by the sonographer to be an ovary. Therefore an expert in gynaecological scanning can subsequently validate the sonographer's judgement by inspecting the image to reach an opinion about the object marked by the callipers: Is it an ovary or not? By repeating this type of validation for images from a set of exams, the expert can independently confirm the key TVS quality control metric of visualization rate (VR). A study that involved such independent confirmation of VR was described in the previous chapter. The aim of the study described in this chapter is to determine a correlation between a novel metric obtained by a software tool and a measure of VR obtained for the same set of exams by an expert in gynaecological scanning.

Methods

The grey scale archived images from the random selection of 1,000 exams were used for the study dataset. The inclusion criteria for these exams were:

- a) Annual TVS exams of women in the ultrasound screening group.
- b) Images stored in the URA.
- c) Performed after 1st January 2008.
- d) Both ovaries reported as visualized by the sonographer.
- e) Both ovaries classified as having normal morphology.

Transabdominal (TAS) only exams and those classified abnormal or unsatisfactory were excluded. Sonographer recorded dimension values (D1, D2, D3) for left ovary (LO) and right ovary (RO) were extracted from the Trial Management System (TMS). This dataset is the same as used for the study reported in Chapter 7.

Expert Review

The expert review was required to provide assessment of the study dataset in terms of left and right ovary visualization for each exam. The tool needed to support this review is described below.

Requirements

The expert review involved one expert who was required to set just one categorical variable in respect of each image to categorise it as visualized and correctly measured, visualized but poorly measured, not visualized, or other.

Design and Implementation

It was decided that the requirements would be best satisfied by creating a spreadsheet of the type described as the first review tool in Chapter 5.

The images from the 1000 exams in the dataset were copied from the URA and used to setup the spreadsheet. These images were then reviewed by an expert in gynaecological scanning. The expert recorded his assessment of the image of the left and right ovary separately using one of the categorical variables. Images which were not of the adnexal region, such as uterus, were marked as not appropriate. In this way each exam was assessed in terms of the visualization of left and right ovary and whether the calliper marks were correctly positioned to measure the ovary.

Quality Control metrics obtained by software analysis of the images

The problem of using a software tool to determine whether the images used to measure the left and right ovary had correctly identified an ovary was expressed in terms of confirming that the shape and size of the object is consistent with the ranges expected for an ovary

It was considered that the functionality needed for this approach could be met by adapting the image workflow described in Chapter 6, so the work focused on how it might be implemented using existing MxPluginForImages components.

Requirements Analysis

Ovaries are described as an olive shaped object (see Chapter 1) so it can be supposed that an object shaped like a potato chip is unlikely to be a normal ovary.

Furthermore, like all other organs in the body, ovary dimensions are expected to lie within a certain range. Therefore, ovary visualization can be judged by simply checking that the dimensions of the callipers are consistent with the expected shape

and size of an ovary. Previous work had provided dimension data for about half the exams in the image dataset (described in Chapter 7) so the requirement analysis was essentially concerned with considering how this data might be used to provide a quality control metric in this regard.

The correct ordering of the D1, D2, D3 dimensions for the left and right ovary for exams in the subset would allow calculation of a novel quality control (QC) metric called Dimensions in Consistent Range (DCR); see Table 9-1. This metric arose from discussions with the UKCTOCS Co-Investigator (an expert in gynaecological scanning) about how the assessment of a static image from a TVS examination might be made more objective in terms of deciding whether or not the sonographer had placed the callipers on an ovary.

Whilst it was recognised that the DCR metric is only a potential first step towards improving the assessment of TVS exams, it was considered to satisfy the need of determining whether the size and shape of the object marked by the sonographer in the image was consistent with it being an ovary. Accordingly it was postulated that the D1, D2, D3 dimensions of an object measured in the image that satisfied the DCR criteria was consistent with the shape and size of an ovary for a post-menopausal woman²²². In order to test this hypothesis, DCR metrics would need to be calculated from dimensions obtained by software analysis of the images associated with the exams in the study dataset. In this way the left and right ovary DCR values for each of these exams might then be tested against the visualization category of the left and right ovary for the exam as assessed by the expert in order to find if any association existed that might justify using DCR rate as a proxy for visualization rate (VR).

Criteria	Notes
D1 \geq 14mm	Ovaries have a minimum dimension
D2/D3 ratio in the range >0.5 and < 1.5	Ovaries have expected shape
Volume (D1*D2*D3*0.523) in range $>0.8\text{cm}^3$ and $<5.0\text{cm}^3$	Ovaries have a range of volumes

Table 9-1: DCR Metric Definition. The metric reflects the size and shape of an ovary as measured by the sonographer against the ranges of values considered likely to be found in normal ovaries of postmenopausal women.

Design and Implementation

The requirements analysis identified the need for some algorithms to process the study dataset. The most important algorithms that needed to be implemented by the functions in the MxPlugin component are described below in terms of their design and implementation.

Algorithms for assessment of object size and shape

The requirements for assessing the object identified in TVS images by the placement of callipers in terms of their size and shape are the same as for the image workflow described in Chapter 7. Therefore it was decided to reuse the same library component (and IPF) without alteration.

The workflow is required to calculate the values of D1, D2 and D3 for the candidate ovaries identified during image processing and store them in the Image Dataset File (IDF) for subsequent analysis using tools like Microsoft Excel. The results of this analysis will then be imported back into the IDF as an additional image property (binary value: ovary – true or false) so it could be used as an input parameter for subsequent image workflow processing. Alternatively, this processing could be performed without such third-party tool analysis by defining an image workflow processing step in the IPF which uses a custom classification algorithm implemented in an MxPlugin component.

Statistical Analysis

Visualization rates were calculated from results of the expert review of all the images associated with the 1,000 exams in the dataset. An ovary was defined as ‘seen’ when the expert reviewer classified the image using any of the categorical variables ‘visualized and correctly measured’ or ‘visualized but poorly measured’. The use of any other categorical variables was defined as ‘not seen’. Visualization rates were calculated for all 1,000 exams in the dataset using the various definitions in Table 9-2 with ‘seen’ corresponding to true (1) and ‘not seen’ corresponding to false (0).

RO or both			LO or both			One or both			Both*		
LO	RO	Exam	LO	RO	Exam	LO	RO	Exam	LO	RO	Exam
0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	1	0	1	1	0	1	1	0	0
0	1	1	0	1	0	0	1	1	0	1	0
1	1	1	1	1	1	1	1	1	1	1	1

LO only			RO only			None		
LO	RO	Exam	LO	RO	Exam	LO	RO	Exam
0	0	0	0	0	0	0	0	1
1	0	1	1	0	0	1	0	0
0	1	0	0	1	1	0	1	0
1	1	0	1	1	0	1	1	0

* visualization of the remaining ovary satisfies the TRUE condition in the case of a woman with a single oophorectomy
 EXAM RESULT
 OVARIES SEEN = 1
 NOT SEEN = 0

Table9-2: Truth tables giving various definitions of Visualization Rates (VR). For example, in the case of ‘Both’ (VR-Both) visualization is TRUE (1) only if the exam left ovary (LO) and right ovary (RO) were seen – i.e. both set as 1. In all other cases visualization is FALSE (0).

The exams in the dataset were processed to create two further subsets:

- 1) The ‘DCR’ subset containing examinations for which software found the exact images used by the sonographer to measure the left and right ovary. Only those examinations where both transverse section (TS) and longitudinal section (LS) for both left and right ovary were present was included in this subset (TSLs-Both).
- 2) The ‘no match’ subset containing the examinations for which software could not find the exact images used by the sonographer to measure the left and right ovary.

Visualization Rates (VR) were calculated from the expert assessment for the exams in both subsets using the definitions in Table 9-2 so that the differences between them could be assessed.

Contingency tables were prepared to allow evaluation of the significance of the difference between the proportions visualized /not visualized of DCR from the software analysis and that as judged by the expert. These proportions were then assessed using Pearson's chi-squared test in order to obtain a P value and relevant

test statistic. The contingency tables were also used to calculate positive and negative agreement using raw agreement indices.

Results

Of the 216,152 annual TVS exams stored in the UKCTOCS URA, 113,092 had been performed after 1/1/08. These exams were classified in terms of outcome as normal (105,176), abnormal (5,097) and unsatisfactory (2,820). The dataset of 1,000 exams was randomly selected from 68,931 of the 105,176 normal exams that had both ovaries reported as 'seen' according to data recorded by the sonographers in the TMS. This dataset had 4,654 images; mean 4.65 images per exam and range 1-15. All the exams were performed by a team of 96 UKCTOCS sonographers (accredited as level I) with only one performing less than 100 exams. The VR definitions given in Table 9-2 were used to analyze the results of the expert review of all images for the 1,000 exams in the dataset and the results are shown in Table 9-3.

VR Definition	Exams with ovarian images identified by software		Exams with ovarian images not identified by software		All exams	
	Count (n=534)	VR %	Count (n=466)	VR %	Count (n=1000)	VR %
RO or Both (right ovary or both ovaries seen)	366	68.5	298	64.0	663	66.3
LO or Both (left ovary or both ovaries seen)	344	64.4	286	61.4	630	63.0
One or Both (left or right ovary seen or both ovaries seen)	430	80.5	362	77.7	792	79.2
Both (both ovaries seen)	280	52.4	222	47.6	502	50.2

Table 9-3: Visualization Rates from expert review by VR definition. Table 9-2 contains a series of definitions for VR and depending upon the definition selected VR varied for the same dataset from 50.2% (VR Both) to 79.2% (VR One or Both).

The software identified a DCR subset of 551 exams (55.1%) containing the exact image used to measure the left and right ovary from the 1,000 exams in the dataset. The information identifying the images used to measure ovaries could not be recovered in the remaining 449 exams (44.9%) for the reasons reported in Chapter 7. The DCR exam subset was further split according to the format of the images giving 534 exams (97%) with one image containing both the transverse section (TS) and longitudinal section (LS) of the left ovary and the other image containing the same sections for the right ovary; defined as TSLS-Both. The images used to measure the ovaries in the remaining 17 exams (3%) had different formats so TS and LS were contained in different images resulting in as many as four images being used to measure ovaries for a given exam. Therefore to simplify processing they were not used in the analysis as stated in the methods section.

Table 9-4 contains the VR results from the expert review of the images in this 534 exam subset as well as for the 466 exams in other subset using the various definitions for VR given in Table 9-2.

Visualization of postmenopausal ovaries on static grey scale archived images	Image used to measure ovary identified by the software		Image not identified by the software	
	Count (n=534)	VR on expert review (%)	Count (n=466)	VR on expert review (%)
Both	280	52.4%	222	47.6%
Left Only	64	12.0%	64	13.7%
Right Only	86	16.1%	76	16.3%
None	104	19.5%	104	22.3%
TOTAL	534	100%	466	100%

Table 9-4: Visualization Rates (VR) from expert review for each subset of the data shows only a small difference between the two sets of values ranging from 4.8% (VR Both) to 0.2% (VR Right Only).

DCR values were calculated for the 534 exam subset using the dimensions gathered by the software from the exact images used by the sonographer to measure the left and right ovary. It was found 91 exams satisfied the criteria for DCR using dimensions obtained from images where the expert reported the left and right ovary as being visualized (true positives), whereas 221 exams failed to satisfy the criteria for DCR and were also not visualized by the expert (true negatives) as shown in Table 9-5. The accuracy of DCR compared to expert evaluation of visualization is

58.4% (CI 95% 54.2, 62.6). The Pearson Chi-Square test applied to the data in Table 9-5 yields a statistical test value of 28.43 with a p value < 0.0001. The raw agreement indices applied to the data give a positive agreement of 0.45 and negative agreement of 0.67.

		Visualization on expert review		Total
		Both ovaries visualized	One or both ovaries not visualized	
DCR	Both ovaries visualized	91	33	124
	One or both ovaries not visualized	189	221	410
Total		280	254	534

Table 9-5: Contingency tables compare DCR to Visualization in 534 exam subset. There is good agreement between the expert assessment of visualization and the software metric DCR in terms of one or both ovaries not visualized, but poor agreements in terms of both ovaries visualized.

Discussion

VR is defined in a number of ways and a review of the literature indicates variation in the definitions used in different studies (Table 9-6). These differences in definition make it difficult to compare scanning quality between studies based on VR results.

The results of our study indicate that depending on the definition of VR used, its value for the same set of exams varies between 79.2% (VR one or both) and 50.2% (VR Both). Therefore there is a need to standardize the definition of VR using one of the truth tables given in Table 9-2. It is recommended that the 'VR Both' definition is used when considering quality control metrics for TVS exams because abnormalities leading to a diagnosis of ovarian cancer may occur in either ovary so it is necessary to visualise them both during the exam. This definition of VR is also consistent with the criteria for exam classification given in the UKCTOCS trial protocol.

Trial or study	Exam numbers	Dates	Definition of Visualization Rate (VR)	Reported Visualization Rate (VR)
UKCTOCS ²	270,035	June 2001-Dec 2011	RO or both	72.7%
			One or both	84.5%
UKCTOCS ³⁰	43,867	June 2001-Aug 2007	RO or both	66.8%
			LO or both	65.5%
PLCO ²²³	102,787	1993-2009	Both	60%
Kentucky ²²⁴	57,214	1987-1999	One or both	79.2%
Kentucky ²²⁵	120,569	1987-2005	One or both	84%
Kentucky ²¹⁶	205,190	1987-2011	One or both	87.6%
Ludovisi et al study ²²⁶	6649	Oct 2008-Sept 2013	RO or both	84.1%
			LO or both	82.4%
Gollub et al study ²²⁷	206	June 1988-Mar1989	Both	49%
			One or both	80%
			None	20%

Table 9-6: Reports of ovary visualization in TVS Scanning shows wide variation in results according to the definition of Visualization Rate (VR) used as specified in Table 9-2.

Visualization Rate Obtained by Expert Review

Applying the ‘VR Both’ definition to the subset of the data for which the exact image used to measure the ovary was identified by the software suggests that almost half of these exams should have been classified as unsatisfactory rather than normal, since the expert reviewing the images did not agree that the object measured by the sonographer was an ovary; see example in Figure 9-2 compared with Figure 9-1.

The study dataset was randomly selected from annual TVS exams of archived images where both ovaries were described as seen and having been seen with morphology described normal, so our results suggest that almost half of such TVS exams may have been unsatisfactory. It needs to be noted that this is the first reported study of expert review of static images of ovaries reported with normal morphology. Additionally, even if the ovary was not visualized, it thought likely that most adnexal abnormalities associated with late stage disease would have been

detected during the detailed examination of the pelvic sidewall due to their heterogeneous morphology.

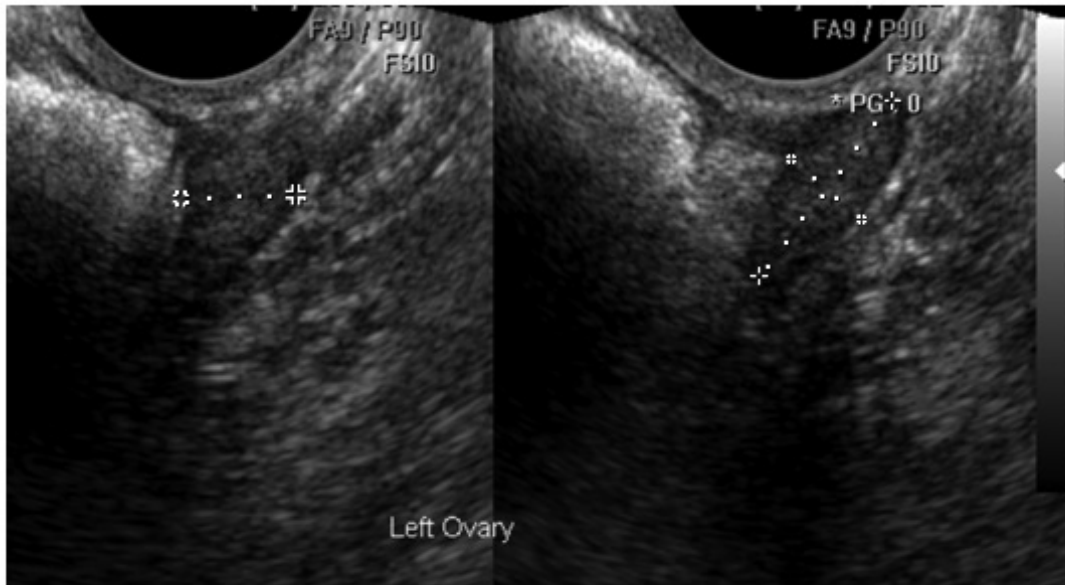


Figure 9-1: Correct measurement of an ovary. This image suggests the sonographer has correctly identified the left ovary and measured it.

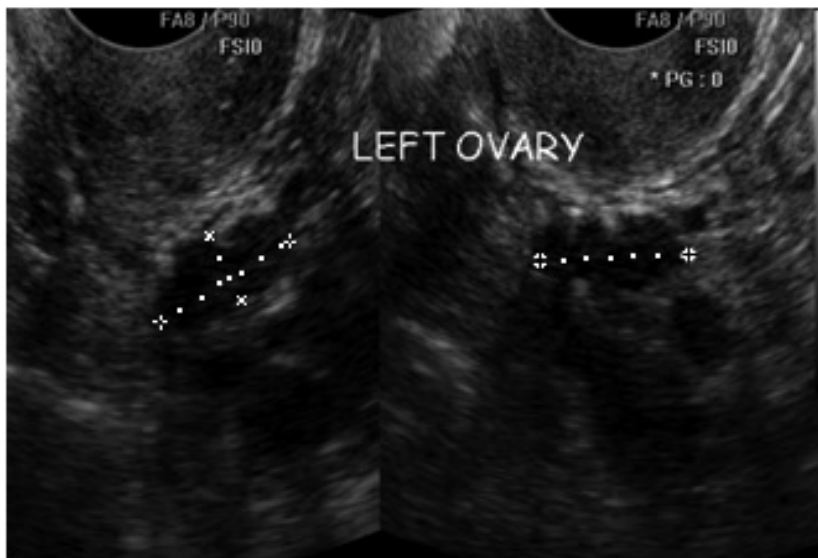


Figure 9-2: Measurement of bowel instead of ovary. This image suggests the sonographer may have measured a section of bowel rather than the ovary. Haustrations of large bowel are suggested by the small pouches that characterize the boundary.

Visualization Rate Obtained by the Software Tool

The results of the Pearson Chi-Squared test (statistical value 28.43, $p < 0.0001$) suggest a high correlation between the DCR metric and the expert review of exam visualization. However, the raw agreement indices of 0.45 and 0.67 for respectively positive and negative agreement suggest that DCR might only be used as a proxy for ovarian non-visualization when auditing a collection of exams. While it would not be possible to signal an unsatisfactory exam using a DCR metric calculated at the time of examination, collecting such metrics for TVS exams performed over a period of time could be useful in assessing levels of unsatisfactory scanning. This raises the possibility of using DCR as a low cost, independent and objective QC metric to drive a Quality Improvement (QI) programme for TVS exams.

When using data from UKCTOCS it is important to obtain the D1, D2, D3 dimensions from the image for as reported in Chapter 7 these dimensions were not always recorded by the TMS in the correct order. However, it should be noted that DCR metrics may be calculated from other data sources without the need to gather dimensions directly from the images if it is known that the D1, D2, D3 ordering is correct.

The data contains a subset of 449 exams for which the exact image used to measure the ovary cannot be identified because:

- a) It was not stored in the URA – i.e. missing (16%).
- b) More than one image for the exam had the same calliper marks – i.e. duplicate (8.6%).
- c) The calliper dimensions were the same for left and right ovary so the images could not be differentiated – i.e. unresolved (16.4%).
- d) There was a software data issue – i.e. process failure (3.9%).

In addition, 17 exams did not have images in the required format so a total of 466 exams were not analyzed. No difference in visualization rates was identified between the expert review of these 466 exams and the 534 where the software identified the images (Table 9-4). Therefore the 534 exams analyzed could be considered representative of the entire dataset.

Other Work

The results of the study reported in Chapter 8 shown a wide variation of visualization rate amongst sonographers and trial centres when confirmed by experts. Although several factors have been reported that could affect visualization (previous hysterectomy, previous tubal ligation, increasing age, unilateral oophorectomy, increased BMI, etc.), Sharma et al² reported only small differences when VR was adjusted for such factors. Therefore these factors were not examined to see if they had impacted in anyway on the expert's review of visualization. This was considered justified given the nature of the study and relatively large size of the dataset (1,000 TVS exams) which had been randomly selected from 68,931 TVS exams where both ovaries were reported to be visualized.

Limitations of Study

It is likely that a degree of inter-observer and intra-observer variability would have impacted the reported results given the subjective nature of grey scale ultrasound and the fact that the review was performed by one expert alone. More robust data may have been obtained by involving three experts in the review to obtain a consensus view of ovary visualization so the results could have been adjusted for inter-observer variability, though advice about the planning and execution of such studies given in Chapter 8 needs to be considered. The review also did not involve repeat assessment of some images at a later date, so intra-observer variability was not measured.

Future Work

The Quality Control (QC) metrics obtained by this study result from analysis of the calliper lines marking the ovary boundaries in the images used to measure ovaries during a TVS exam. They allow verification that the size and shape of the object measured was consistent with that of an ovary. However, more conclusive evidence would be provided by assessment of this object's boundary and content. The prototyping work performed using active contour models as described in the software design section shows potential to segment the ovary and thereby overcome a significant barrier to undertaking such assessment.

Future work would involve the development of further prototypes to assess the feasibility of automating the segmentation of ovaries from TVS images of post-menopausal women. However, a precondition for such work is the means to verify that the software had correctly segmented ovaries from TVS images. This suggests the need to create a collection of images containing ovaries whose boundaries had been marked by an expert in gynaecological scanning using an annotation tool such as ImageCurator, as described in Chapter 5. This would allow experiments to be conducted using the image workflow tool in order to select the algorithms needed for segmentation and the optimisation of their parameter settings as proposed in Chapter 6. The success of such work in allowing the reliable segmentation of ovaries from a large number of exams would allow further processing to obtain properties for ovaries and non-ovaries such as texture and boundary composition. These properties could be used to train a classifier able to distinguish between those images containing an ovary and those which do not. The work of building such a classifier and integrating it with the workflow tool is comparatively straight forward, though some effort would be needed in this regard particularly in terms of testing and validation. In this way useful tool might developed that is able to obtain at low cost advanced QC metrics from individual TVS examinations that were both objective and independent.

Summary

This chapter has identified the need to standardize the definition of Visualization Rate (VR) in terms of the options presented in Table 9-2. This would facilitate comparisons between different studies for the purposes of benchmarking sonographer or trial performance. In this regard the definition VR-Both has advantages when VR is being used as quality control metric as it requires both ovaries to be visualized. This is important when TVS scanning is used as a screen for ovarian cancer as abnormalities associated with the disease may occur in either ovary.

The previous chapter provided evidence about the reliability of ovary visualization assessments made by the examining sonographer during TVS scanning. It described an audit involving a team of nine experts in gynaecological scanning reviewing images from 223 TVS normal exams performed by 7 sonographers. The audit suggested that more than a third of the TVS exams performed by a group of 5 sonographers which had been classified as showing ovaries with normal morphology should be reclassified as unsatisfactory. This was because in the opinion of the experts the images used to measure the ovaries did not support the claim that both left and right ovary had been seen. The study described in this chapter adds to this evidence by reporting the results of a review by a single expert of images from 1,000 randomly selected TVS normal exams performed by 96 sonographers. However, in this case it was suggested that almost half of exams classified as showing ovaries with normal morphology should be reclassified as unsatisfactory for the same reason.

It can be concluded that QC metrics based on self-assessment of ovary visualization by the examining sonographer are problematic. However, a solution is suggested in terms of using the workflow tool described in Chapter 6 to automate the analysis of the images used to measure ovaries in such exams. The study described in this chapter demonstrated that this tool was able to provide novel QC metrics for TVS exams which might be used as a proxy for VR for collections of exams. Furthermore the prototyping work completed as part of this study suggests the feasibility of enhancing the tool in order to provide a classifier able to distinguish automatically between the images that contain an ovary and those which do not in the case of

individual exams. Therefore the use of such a tool might allow the routine generation of QC metrics for TVS scanning which are independent and objective in order to drive improvement in TVS scanning. Furthermore, as shown by the study described in Chapter 7, such metrics may be obtained reliably and at low cost.

Conclusion

Chapter 10: Thesis Conclusion

Introduction

UKCTOCS is the first large-scale trial to have routinely archived images from TVS examinations and the use of such images to retrospectively assess the quality of scanning in the case of morphologically normal ovaries has not previously been attempted. UKCTOCS has also undertaken pioneering work in other areas of TVS scanning. For example, the training and accreditation of its sonographers has set high standards which others are now following. Therefore the work of UKCTOCS can be viewed as part of the learning process given there is open and honest discussion about the lessons learnt and how improvement might be made.

The thesis conclusion starts providing an answer to the research question stated in Chapter 1. It then considers how the quality control metrics obtained from the type of tools discussed in previous chapters might be used in the future large-scale delivery of TVS scanning as well as the wider requirements for achieving meaningful quality goals in respect of such work. The conclusion ends by considering the impact of the thesis project and the future work required to progress the research it has initiated.

Answer to the research question

The research question stated in Chapter 1 asked: Can software tools be developed to collect independent and objective QC metrics for TVS exams by routinely analysing the images generated for ovary measurement? The tools and research performed by the thesis project suggest that this question can be answered in the affirmative.

The development of the image processing workflow and its associated MxPlugin components has produced a software tool that is able to reliably generate QC metrics from the images found to have been used by the sonographers to measure ovaries during a TVS exam. The study described in Chapter 9 shows high correlation between the DCR metric produced by this tool and the results of an expert review in the case of images from a collection of exams. This makes a strong case for developing software tools so that independent and objective quality control metrics can be gathered at low cost from the images used by the sonographers to measure ovaries during TVS exams. However, there is no point in creating quality control metrics without having a process that allows use them to be used to drive quality improvement.

Delivery of High Quality TVS Scanning

The delivery of high quality TVS scanning in any future trial or national programme to screen women for ovarian cancer using transvaginal sonography requires not just better QC metrics, but also the establishment of a management team capable of using them to drive quality improvement. Such a team needs to introduce independent oversight and change management capability in addition to implementing the type of quality processes that have provided so much benefit to other areas of healthcare.

Audit

The SPIRIT 2013 Statement²³⁵ aims to improve the quality of clinical trial protocols by providing a checklist of items that they need to address. It is endorsed²³⁶ by many of the world's leading journals, regulators, funders and research organisations including The Lancet, British Medical Journal, Medical Research Council Clinical Trials Unit (UK) and University College London Clinical Trials Unit.

Item 23 of the SPIRIT checklist is concerned with the auditing requirements for clinical trials. Chan et al²³⁷ states “*[auditing] is distinct from routine day-to-day measures to promote data quality... the procedures [for audit] and anticipated frequency should be outlined in the protocol, including a description of the personnel involved and their degree of independence from the trial investigators and sponsor.*” The statement suggests that routine audits of the trial processes should be performed by people who are independent of the organisation concerned with the trial’s operation and administration. The benefit of such regular auditing is illustrated by the work described in Chapter 8 which identified a problem concerning the reliability of a key TVS scanning quality metric; ovary visualization rate. It might have possible to address this issue, or mitigate its impact, had the audit been conducted at an early stage of the trial.

Quality Plan

The creation of a quality plan describing its Quality Assurance (QA) processes and associated Quality Control (QC) is consistent with good governance and will also help any trial to comply with item 23 of the SPIRIT statement by documenting the basis of verification by audit. Work on the development of the quality plan needs to start at the inception of the work and then continue iteratively as the work progresses; an approach taken by the Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening trial, as described by Weissfeld et al¹²⁴.

The information provided in Chapter 3 may help in the formulation of an approach to quality management. It explains QA as a cycle involving setting standards, measuring practice, comparing practice against standards and then making improvements in terms of both practice and standards. By way of example, consider how QC might be introduced into a process for confirming the correct completion of a particular form by asking someone who isn’t involved in its day-to-day processing to check that in a random sample certain fields contain values within ranges specified by quality control standards. It is clear that without defining standards and measuring compliance in a systematic way it would be difficult to obtain quantitative information about defect rates which could drive improvement in both initial form completion and its subsequent checking.

Quality Governance Framework

The creation and maintenance of an effective quality plan requires the involvement of people with knowledge, understanding and experience in such matters. Their work should also be guided by the creation of a quality governance structure such as proposed by the National Quality Board (NQB)²³⁸ as mentioned in Chapter 3. The framework suggested by the NQB describes the governance of quality for NHS Provider Boards, but it might be adapted in terms of:

- **Strategy** – the aims and objectives of the quality plan should be clearly stated and its implementation should evolve as the work progresses with particular attention to mitigating the risks of poor quality. The delivery of the quality plan's objectives should form a key part of any trial's work for there is little value in research that produced poor quality data. Similarly public confidence in any national screening programme would be eroded by high levels of incorrect screening results, positive or negative, resulting from poor quality tests for ovarian cancer.
- **Capabilities and culture** – a management group should be formed with the leadership, skills and knowledge necessary to formulate and deliver the quality strategy as well as to promote a quality focused culture amongst all the people participating in the work. It is particularly important for the group to attract people with expertise in quality management with equivalent seniority to clinical specialists in order to achieve a degree of balance and diversity.
- **Processes and structure** – there should be clear roles and accountability with regard to the delivery of the quality plan with well understood procedures for managing quality performance as well as escalating and resolving issues. Quality Assurance (QA) processes should be defined to ensure the requirements of the quality plan are satisfied in terms of specific activities being undertaken. The operation of these processes should be subject to routine independent audit.
- **Measurement** – independent and objective Quality Control (QC) metrics should be collected and analyzed as defined by the QA processes so the performance of specific activities can be compared to set standards in order

to drive improvement. The management team should first focus on controlling quality and then progress to quality improvement as suggested in Chapter 3.

People need to be recruited with the expertise to create an effective quality strategy and put in place the necessary management structure so that quality improvement becomes a core activity supported by appropriate QA processes within a culture where quality is highly valued at all levels.

TVS scanning used as a screen for ovarian cancer

There is a difference between TVS scanning being used as a screen rather than a diagnostic for ovarian cancer as discussed in Chapter 1, Development of Ovarian Malignancy. Sonographers usually perform a TVS scan as a diagnostic after a woman has reported symptoms associated with ovarian cancer. The presence of such symptoms is associated with late development of the disease (stage III or IV) when the tumour is typically large and considered difficult to miss during TVS examination. However, a successful screening programme for ovarian cancer would depend upon the identification of tumours at a much earlier stage of their development when they might be only slightly larger than 1cm in diameter, as suggested by Brown and Palmer²⁵. Such tumours are close to the limit of what might be detected by TVS scanning using current technology. Furthermore, when the examination is given as a screen for ovarian cancer it needs to be approached in a way that supports a QA process. Therefore TVS scanning needs to be performed differently when it is used as a screen rather than a diagnostic for ovarian cancer. This is not just in terms of what it is attempting to detect, but also in terms of it being a structured process rather than a general investigation of the woman's adnexa. In this later respect it may be possible to subject parts of TVS examinations to the type of quality control performed for nuchal translucency measurements used as screen for Down's syndrome, as described in Chapter 3. For example, the ovary measurements made by each sonographer could be aggregated and compared to standard ranges to detect those whose results suggest they are outliers in respect of their peers.

When performing TVS scanning as a screen for ovarian cancer sonographers need to be given specific training so they know what they are looking for during scanning, particularly if their previous experience only arises from ovarian cancer diagnosis by the identification of the substantial tumours associated with the onset of symptoms in late stage disease – i.e. tumours considered difficult to miss during transvaginal sonography. Whilst the difference between a screen and diagnostic may appear to be only a subtle change of viewpoint, helping sonographers to understand the real objective of their work has clear importance in terms of them finding normal ovaries and any small tumours they may contain. It will be necessary to develop new protocols in this regard as there is no published guidance and scant knowledge about what is required, though the ability of sonographers to visualise both ovaries is obviously a prerequisite to finding small tumours. The experience gained by UKCTOCS in delivering more than 300,000 TVS scans as an ovarian cancer screen should be influential in the development of such protocols and it is anticipated that scientific papers will be published by the UKCTOCS group in due course to advance knowledge in this area.

Quality Control of Ultrasound Equipment

Ultrasound machines require regular inspection and maintenance even though modern electronics has reduced the need for periodic recalibration. Chapter 3 described the approach to the quality assurance of ultrasound equipment adopted as suggested by BMUS¹⁰² and others. This involves routine technical quality checks to identify problems such as scan line drop-out and the grayscale integrity of display devices as well as changes in sensitivity and noise as detected using calibrated test objects. Sonographers also need to perform frequent checks of their equipment. These include infection control, scanner damage, basic functional tests as well as more advanced operational tests.

Any large-scale TVS scanning programme needs to implement a quality assurance process for ultrasound equipment to ensure it is safe for use and maintained in an operational state that is satisfactory for the type of examinations being performed. This will involve creating service contracts with equipment vendors, as was done by UKCTOCS. It will also involve training sonographers and documenting the quality control checks performed on the equipment in order to provide evidence that it was

safe and suitable for use in TVS scanning. In this way variation in scanning quality arising from equipment problems can be minimised which in turn will allow better quality data to be recorded.

Training and Accreditation of Sonographers

It is difficult to perform a TVS scan in a standard and repeatable way, even when it is being used as a screen for ovarian cancer rather than as a diagnostic. Therefore the type of scanning performed by UKCTOCS is highly dependent on employing sonographers with appropriate training, skills and experience. This was recognised by the UKCTOCS Ultrasound Management Committee in 2006 and resulted in the appointment of the National Lead Sonographer who helped instigate the training and accreditation of more than a hundred sonographers who were performing TVS examinations for the trial; see Chapter 3. The ground-breaking work accomplished by UKCTOCS in this regard has been recognised by the British Society for Gynaecological Images by the adoption of the Standard Operating Procedures as guidance for its good practice notes and the modification of the accreditation process into its continuing professional development programme.

The implementation of a training and accreditation process for sonographers and the provision on-going support as well as professional development helps ensure consistency of scanning. Attention to such matters is particularly important in any large-scale delivery of TVS scanning as it is fundamental to achieving dependable quality.

Provision of Information Technology Systems

Information technology (IT) has advanced considerably since UKCTOCS was planned in the late 1990s. The products and services commonly available in 2016 would have greatly facilitated the implementation of the trial, as would some of the novel tools described in this thesis. Therefore consideration needs to be given to how these tools and other recent IT developments might benefit the future delivery of large-scale TVS scanning within the scope of quality management.

The specification and procurement of IT systems requires considerable experience and expertise. Therefore significant benefit may result from employing a consultant

to provide independent oversight of vendors to ensure that they deliver systems that satisfy the true requirements at the time they are needed and with an acceptable level of quality. Such professional oversight may also reduce the costs associated with developing such systems and help set realistic budgets. In this respect it is worthwhile reflecting on the experiences of others. The budget for IT systems set by UKCTOCS at grant application was approximately £750,000, with an allowance for maintenance of about £100,000 per annum. However, almost the entire budget was allocated to a Trial Management System (TMS) as only £2,000 was spent on the development of the Ultrasound Record Archive (URA) as discussed in Chapter 4.

Although the UKCTOCS IT systems operated successfully throughout the trial and delivered the data needed to report its main findings, it is suggested that a different approach to procurement may have allowed the delivery of additional functionality with significant value, particularly in terms of allowing better quality control and auditing. In this regard the systems developed by the Prostate, Lung, Colorectal and Ovarian (PLCO) trial appears to have been superior in a number of areas, as described in Chapter 3. However, this suggests the PLCO trial had a significantly larger budget for IT systems than was the case for UKCTOCS.

Collection of QC Metrics

Consideration needs to be given to the way in which the quality of images from TVS scanning could be monitored by using a computer system to collect QC metrics as discussed below.

Ability to capture images of an ovary

Scanning quality can be judged in terms of the ability of a sonographer to capture a static image of an ovary that can be recognised as such by subsequent review by experts in gynaecological scanning. The results of the audit described in Chapter 8 suggest that some sonographers are better than others in this regard. It also found that some sonographers can visualize both ovaries in most of the women they examine, even those for whom visualization is considered challenging due to factors like age, high BMI, etc.

UKCTOCS required that sonographer record static images of ovaries in transverse and longitudinal section (TS, LS) showing the calliper marks used to measure them

in the three dimensions D1, D2, D3. These callipers marks greatly facilitate subsequent review as the expert needs only to make a decision about whether or not the callipers mark an ovary rather than having to undertake the more difficult task of identify the existence of an ovary somewhere in the image. Two approaches to this problem are discussed in Chapter 9; confirming its size and shape are consistent with an ovary (DCR metric), or assessing whether its boundary and content are ovarian in nature.

The workflow software of the type described in Chapter 6 could be used to automate the analysis of images from TVS exams in order to generate QC metrics reflecting the ability of the sonographer to measure an object that is likely to be an ovary as outlined in Chapter 9.

Placement of calliper marks

The problem concerning the sonographer placing callipers on objects that are not ovaries has been considered above. However, there is also the related issue of sonographers failing to correctly position calliper marks on the boundary of an ovary, or measuring dimensions in axes which are not at right-angles to each other. The consequence of such quality failings is that ovary dimensions and the resultant ovary volume calculation may yield inaccurate data. This issue has particular importance given the research being undertaken in the UKCTOCS group which is attempting to discover associations between small changes in ovary volume and subsequent development of disease.

The study described in Chapter 7 describes how the image workflow tool was adapted to measure the angle between D1 and D2 in order to produce the QC metric D2O; a binary value reflecting whether the angle was within an acceptable range. Further development of workflow to allow automatic ovary boundary detection, as described in Chapter 9, would allow the collection of a metric reflecting whether the calliper was placed within an acceptable distance of the boundary.

Dimensions Assigned Correctly (DAC)

Scientific knowledge about the dimensions and position of a normal ovary is limited as most existing study is focused on ovaries with morphological abnormalities. In this respect the archive of images from more than 200,000 TVS exams together with

the information held in the Trial Management System has much potential to provide better understanding. For example, whilst it is known that the longest dimension (D1) is different to the other two dimensions (D2, D3) it is not known if there is any biological significance to D2 in respect of D3. The TVS scanning performed by UKCTOCS required D1 and D2 to be measured at 90° to each other in the same plane which allows D2 and D3 to be differentiated, but it is not known if there is any meaning to the difference between them. The question is complicated because the ovary may move during examination, though it is held by ligaments and tissue which suggests that its orientation is not entirely variable.

The binary DAC metric collected for the study described in Chapter 7 reflects whether or not the sonographer measured D1 and D2 in the same plane, and if so, whether the three dimensions were recorded in the correct fields in the TMS. Therefore whilst the value of the DAC metric in terms of quality control is limited, it could be used to select exams with ovary dimensions measured in a consistent way and consequently have utility in a study attempting to determine the ranges of dimensions that exist in normal ovaries.

Static Image Quality

The settings of the ultrasound machine influence the quality of the image as their values are displayed in the image itself; see Figure 2-7. Therefore the quality of images used to measure the ovaries can be judged in terms of:

- **Depth of field** set so the ovary is shown within the middle third of the image
- **Focus** set so the ovary is in the focal zone
- **Power** set so the image exposure (brightness) is optimal for showing the ovary
- **Dynamic range** set so the ovary has optimal contrast – i.e. optimal range of greyscale values used to display the ovary and its features
- **Gain** set to avoid cut-off of weak reflections (too low) or strong reflections (too high)
- **Time Gain Compensation** set to have uniform brightness in the Y axis – i.e. suitable compensation applied for attenuation of the signal due to depth

- **Zoom** set to display the ovary at an optimum size for display and measurement. If the zoom is too low then the ovary cannot be measured accurately, but if it is set too high then other features like iliac vessels that can help with its identification may be lost.

It would be possible to assess image quality by comparing the values used for a given TVS examination against normal ranges for each of these settings, possibly allowing compensation for such factors as the volunteer's age and body mass index. However, these values are not recorded in the DICOM headers of UKCTOCS images. Therefore automating the assessment of image quality against normal ranges of these settings would require the reading of values from the image using Optical Character Recognition (OCR) software.

The literature does not contain any reports about attempts to perform quality control of TVS images in respect of specific machine settings. However, subjective measures of image quality by experts in gynaecological scanning have been reported such are those made as part of the study described in Chapter 8. In this case the appropriateness of the settings used by the sonographer for a given TVS exams were aggregated in terms of an ordinal variable for image quality ranked between 1 (bad) and 5 (excellent).

The cost and challenges of performing audits of TVS exams by asking experts to review the images used to measure the ovaries make subjective assessment of image quality problematic, so it is preferable to develop a system able to automate the gathering of QC metrics for image quality based on the machine settings used by the sonographer combined with details about the women being scanned. It would be easy to collect the necessary data if the machine settings were recorded in the DICOM header, but more difficult if OCR software was needed to obtain values from the image as would be the case for UKCTOCS images.

Conclusion

The planning and implementation of any future large-scale trial like UKCTOCS is a significant undertaking and will probably involve ground-break work in many areas. It will also involve people working outside their field of expertise. For example, a clinician might undertake the setting-up of a quality governance framework, or a

researcher may become involved in the complex and specialised work associated with the specification of a computer system. Therefore it is important for a wide-range of experts to be engaged in the planning, implementation and management of the trial. Future trials promise to be more complex rather than less, so the need to create a diverse team is likely to become more pronounced.

The work done by the UKCTOCS group on the Quality Control (QC) of TVS scanning seeks to improve understanding of the challenges associated with performing such examinations as a screen for ovarian cancer on a large-scale and at multiple centres. In this respect it is important to obtain a clear understanding of what the examination is attempting to achieve and then define effective quality control standards to ensure the successful delivery of this goal. Information about such goals and targets needs to be effectively communicated to the sonographers who are performing the examination together with the provision of appropriate training, support and equipment so they can do their work to the required standard. In addition, quality assurance process must be established in accordance with an overarching quality plan. The quality control metrics used to drive such QA processes need to be obtained independently of the people doing the work. The metrics also need to be objective and obtainable at low cost in order to ensure routine quality control is carried-out. The correct operation of the quality processes should be confirmed by regular audits conducted by people who are independent of trial investigators, as suggested by Chan et al²³⁷. In this way people will have confidence in the quality control metrics and the way they are used.

It is crucially important to identify a suitable goal for quality improvement that can be shared by everyone involved in the process so that good quality control metrics are not just collected and analyzed according to a quality process, but also used to drive change in the organisation to achieve objectives with meaningful impact such as delivering better patient care at low cost. There is simply no point in gathering quality metrics if they are not going to be used to attain such goals.

Impact of the thesis project

In addition to the potential contribution of the thesis and three journal papers to provide a understanding of quality control for TVS examinations, the thesis project has also had impact in other ways.

Enhancing the value of the URA

It is recognised that the images from TVS examinations archived by UKCTOCS provide a unique resource. This is the first time that such images have been retained by a large-scale clinical trial involving TVS scanning so the archive has considerable potential in terms of providing information for future research as well as for providing images to assist with the training of sonographers. The value of the URA is enhanced by a facility to create datasets of images for a collection of TVS exams that can be processed by the workflow tool. Existing work allows identification of the exact image used by the sonographer to measure the ovary and endometrium for a given exam. This allows further research based on the selection of such images from others crated during scanning.

Useful tools for other researchers

Further development of the language workbench, domain specific language and generic image processing functions will allow image research to be undertaken by researchers who lack experience and skills in computer programming.

Future Work

Further development of the tools created for the thesis project is planned with the aim of evolving them into high quality products that can be installed and used by third-parties without significant help and support. In this respect it is hoped that the open-source projects will attract the support necessary to implement a formal development, test and release process as well as the user-base needed to drive innovation. Specific areas of development are considered below.

Open-source projects

The following open-source projects hosted by Codeplex have been created to share the work of the thesis project; OSPACS project for the development of the URA,

MxPlugin project for development of C++ components providing binary compatibility, MxPluginForImages project for MxPlugin components that support the interfaces required by the image workflow. Further open-source projects are also being considered to encourage the development and use of the other tools developed for the thesis project.

Production version of the ImageExpter

The existing version of ImageExpter was developed as a prototype for government funded project (InnovateUK). Work needs to be undertaken to build a production version of ImageExpter so it can be used by researchers to perform a variety of image processing tasks without writing computer code.

Segmentation of ovaries from TVS images

Continuation of the prototyping work described in Chapter 9 to create an MxPlugin component for the image workflow able to automate the process of segmenting an ovary from a TVS image, though this requires an expert in gynaecological scanning to annotate a small dataset of images with the locations of ovary boundaries.

Classifier to discriminate between images containing ovaries

A large dataset has been created which identifies images that have callipers marking an ovary and those that do not (see Chapter 9). This dataset in conjunction with a tool able to segment ovaries from TVS images (above) could be used to train a classifier capable of discriminating between images on the basis of their callipers marking the location of an ovary.

Claims as to novelty and contribution to knowledge

The thesis presents novel work and contributes to scientific knowledge in the following key respects:

- Presents the first studies to suggest the gross under-reporting of unsatisfactory UKCTOCS TVS scans – the 2.5% of scans reported as unsatisfactory did not include 30%-50% scans which had been classified as normal, but were actually unsatisfactory as both ovaries had not been seen.
- Presents the first studies to suggest that Visualisation Rate (VR) is not a reliable quality control metric when self-assessed by the examining sonographer – VR is grossly over-reported by sonographers in exams classified as normal.
- Presents the first study to suggest that a few sonographers are able to visualise both ovaries in almost all the post-menopausal women they examine, though most sonographers do not have this high level of skill.
- Describes a novel metric for TVS quality control which correlates with VR suggests the potential to drive quality improvement by automating the collection of metrics from TVS images – such work has not previously been reported.
- Describes a novel algorithm for detecting the location, length and angle between calliper marks in TVS images. No algorithm having such capabilities has been previously reported.
- Describes a novel workflow for processing thousands of TVS images – there has been no previous reports of a system having such a capability. The general nature of the system gives potential for future research involving the >200,000 TVS exams stored in the Ultrasound Record Archive (URA).
- Describes the recovery of key information from images collected from a TVS scan – the exact images used to measure ovary were identified in 55% of exams facilitating future studies based on the ovarian images stored in the URA. This work has not previously been attempted.
- Describes novel software tools to allow curation of images in URA which allows the generation of image datasets containing specific identified

features, like ovaries with inclusion cysts. This will help the URA to deliver on its potential as a unique resource for the study of TVS images by the scientific community.

Summary

Whilst any criticism of United Kingdom Collaborative Trial of Ovarian Cancer Screening (UKCTOCS) is beyond the scope of the thesis, there is a need to disseminate the understanding gained for the benefit of any future work in this area. There is also a need for the thesis to contribute to the body of scientific knowledge in the area of quality control for transvaginal sonographic (TVS) scanning used as a screen for ovarian cancer. For these reasons it is important to discuss the lessons learnt from the work of UKCTOCS.

The thesis project has revealed a quality issue with the TVS scanning performed by UKCTOCS sonographers. In a review by an expert in gynaecological scanning of images used to measure ovaries from 1,000 exams recorded by the sonographer as both ovaries visualised and having normal morphology it was found that almost half may not have been satisfactory; see Chapter 9. These results support the findings of a previous audit of seven sonographers reporting high ovary Visualization Rates (VR) by nine such experts with similar exam selection criteria which suggested that more than a third of the exams performed by a group of five sonographers were similarly not satisfactory; see Chapter 8.

It is understood that the scientific papers arising from the thesis work are the first to provide independent evidence of a quality problem with TVS examinations. These findings may have implications beyond UKCTOCS for the scanning protocol was based on standard practice in the UK National Health Service (NHS) and the sonographers are part of the group that perform over a million such examinations each year for the NHS²³⁹ at significant cost^a. Certainly, should the NHS decide to implement a national screening programme for ovarian cancer then the quality control of TVS scanning will have great significance even if it is used as a secondary level diagnostic rather than as a primary screen. In this regard the type of software tools developed for the thesis project may help to provide at low cost the sort of

^a NHS England Diagnostic Imaging Dataset (DID) 2013-14 reported almost 1.2 million abdominal or pelvic ultrasound examinations were performed mostly to diagnose ovarian cancer in post-menopausal women in 2013-14. A general ultrasound imaging cost of £65 is given by University College London Hospital (UCLH) in the 2014-15 Provider to Provider services tariff. Therefore the annual cost of such scanning in the NHS may be estimated as £78 million. This suggests that had the NHS delivered the 300,027 TVS scans performed by UKCTOCS then the cost would have been approximately £20 million in terms the full direct and indirect charges (at 2014-15 prices).

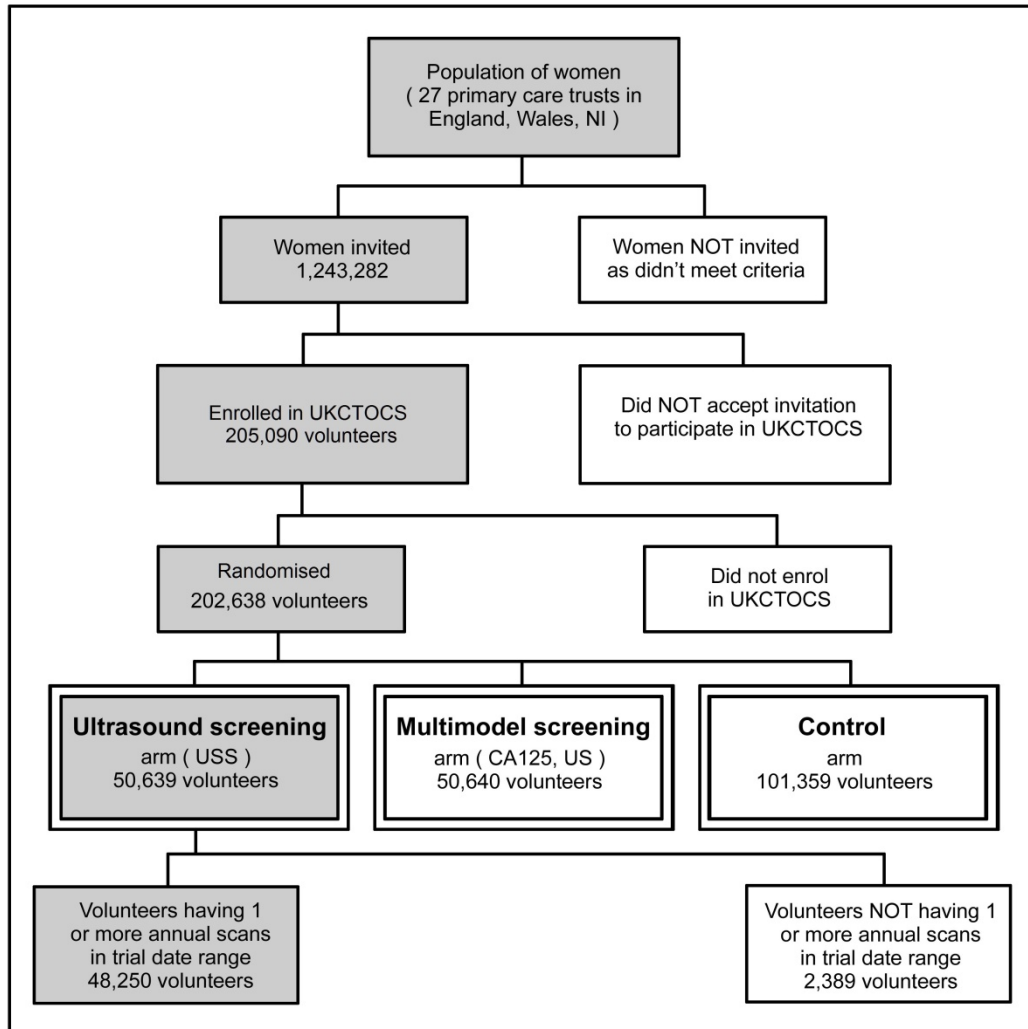
independent and objective quality control metrics needed to implement a Quality Improvement (QI) programme as described in Chapter 3.

The potential of a QI programme to raise standards in TVS scanning is suggested by the findings of the previously mentioned audit which in addition to identifying five sonographers whose self-assessed VR was judged unreliable also identified two sonographers whose VR seemed to be substantiated. This suggests that it is possible for sonographers to visualise both ovaries when scanning menopausal women regardless of factors such as old age and high BMI which are considered to make ovary visualization challenging. Therefore the introduction of a QI programme driven by reliable quality control metrics for individual sonographers might help with the targeting of the training and support necessary to allow most sonographers to achieve similar levels of performance. In this way the substantial costs of TVS scanning might be cut by reducing the number of repeat scans whilst also improving patient care by ensuring more women diagnosed with ovarian cancer had abnormalities in their ovaries detected during their first TVS scan and so were able to receive treatment at an earlier stage of the disease when the prognosis is better.

Appendices

Appendix A: Source of Data for Research Studies

AR 21/1/2016 rev 1.1



Trial inclusion criteria :

Women aged 50-74 years at enrolment who were postmenopausal defined as either :

- a) amenorrhoea >1 year following natural menopause or hysterectomy
- b) received HRT >1 year started for menopause symptoms

Trial exclusion criteria :

- . Previous ovarian malignancy
- . History of bilateral oophorectomy
- . Active non-ovarian malignancy (documented persistent or disease or treatment received in last year)
- . Increased risk of familial ovarian cancer
- . Participation in other ovarian cancer screening trials

Figure A-1: Selection of Volunteers for TVS Scans in UKCTOCS

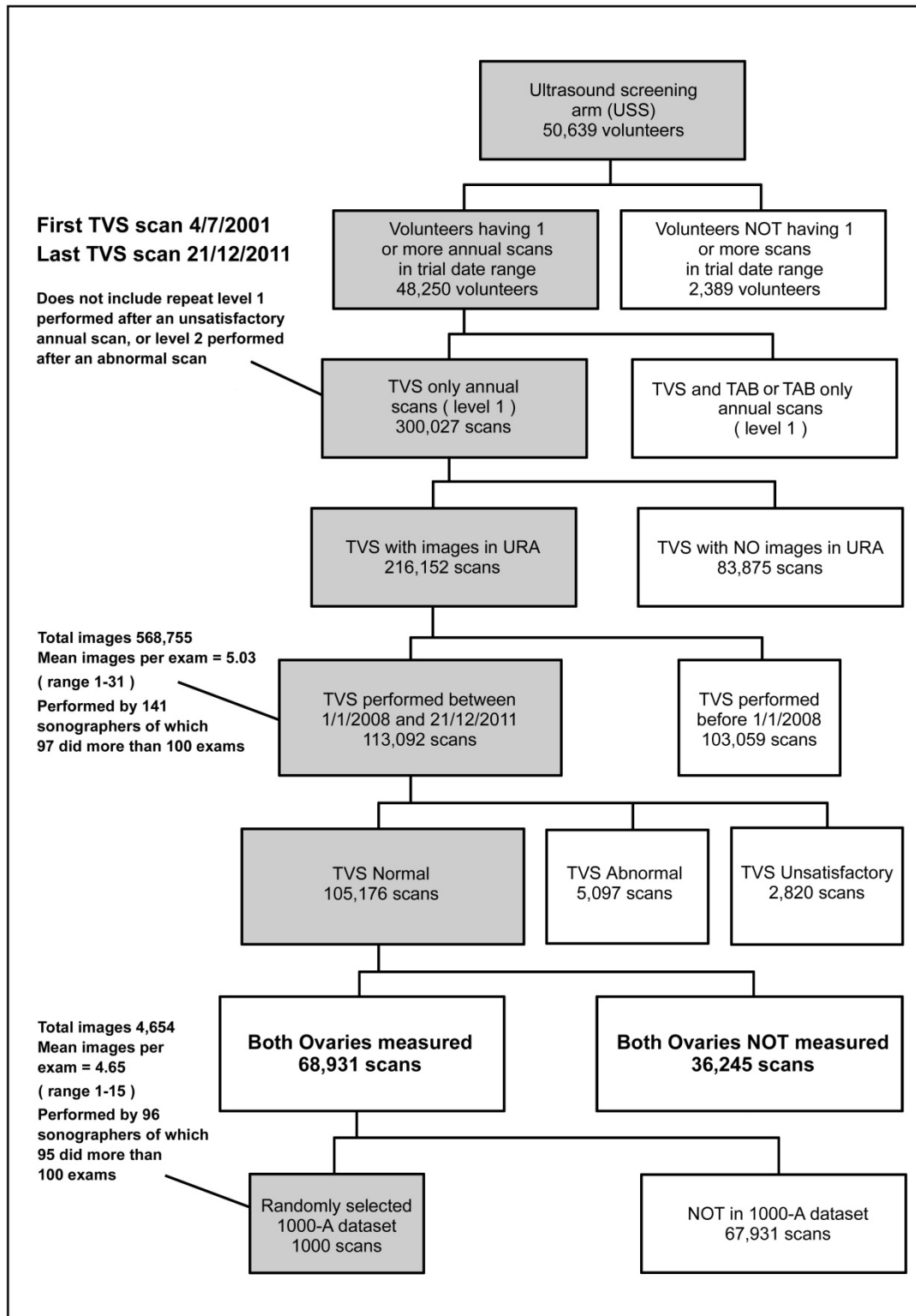


Figure A-2: Source of TVS scan images used for thesis research studies

Appendix B – Extracts from the UKCTOCS Protocol

The following parts of the UKCTOCS Trial Protocol (version 6, Dec 2010) are relevant to the ultrasound arm of the trial and have been copied below verbatim with kind permission of the copyright holder, University College London.

Trial Centres (pages 7,8)

“The trial is conducted through 13 trial centres (TC) in England, Wales and Northern Ireland which are mainly located in NHS Trusts. Each centre has a dedicated trial team consisting of a research nurse, phlebotomist, clerk and ultrasonographers led by a consultant clinician (Lead Researcher). The trial centres undertake recruitment of the participants and performance of the screening tests. The teams arrange clinical evaluation of women confirmed to have abnormalities on screening and referral to the appropriate NHS clinic if surgery is appropriate”.

Interventions (page 16)

“The screening strategies utilise two screening tests – serum CA125 interpreted the Risk of Ovarian Cancer (ROC algorithm and transvaginal ultrasound (TVS) of the ovaries”.

Screening Tests (pages 16-18)

“For the measurement of serum CA125, a blood sample is taken in Greiner gel tubes (8mL gel separation serum tubes; Greiner IO-One 455071, Stonehouse, UK) at the trial centre and transported overnight at ambient temperature to the central laboratory. The blood is centrifuged at 4,000 rpm for 10 minutes and the serum separated. Excess serum is aliquoted and stored. Serum CA125 levels are determined by commercial enzyme immunoassay on the Roche EIA Elecsys 2010 system. All blood samples received >56 hours after venepuncture are discarded and repeat samples requested.

The second test used is transvaginal ultrasound (TVS). Where this is not acceptable to a volunteer, transabdominal ultrasonography is performed. The scans are performed by two levels of staff. Annual and repeat scans following an unsatisfactory annual scan are performed by Level I sonographers who are certified ultrasonographers, or trained midwives or doctors with experience in gynaecological especially transvaginal scanning. Level II scans are performed when an abnormality is detected on the annual screen. Level II sonographers are experienced gynaecologists or radiologists or senior ultrasonographers (usually at superintendent grade in the NHS) with particular expertise in transvaginal ultrasonography. Majority of the scans were performed on a dedicated trial ultrasound machine.

Ovarian morphology and dimensions are assessed and volume determined using the formula for an ovoid ($d1 \times d2 \times d3 \times 0.532$). Ovarian morphology is classified as:

- (1) Normal: if the ovary is of uniform hypoechogenicity and smooth outline with or without a single inclusion cyst or spots of calcifications. To classify as normal the inclusion cyst must be single, less than 10mms and should not distort the outline of the ovary.*
- (2) Simple cyst: A single, thin walled, anechoic cyst with no septa or papillary projections Is detected*
- (3) Complex: All non-uniform ovarian echogenicity excluding single simple cyst. If there are more than one cyst even if these are without septae or papillations and have thin wall with regular internal outline, ovarian morphology is classified as complex.*
- (4) Complex unchanged: Where the morphology is complex as described above but it has been noted and reviewed previously and has remained unchanged in appearance and size on follow up.*

Detailed description of all features - the number and size of cysts, wall regularity, presence and thickness of septae, size of papillations and echogenicity of the fluid contents will be recorded (Appendix O). Definitions of ultrasound features and classification of cysts follow guidelines published by the International Ovarian Tumour Analysis (IOTA) group 9.

When ovaries are not visualised, ultrasonographer specify if a good view of iliac vessels is obtained or if the view is obscured (poor) due to bowel, fibroids, pelvic varicosities or other reasons.

Based on visualisation and morphology of the two ovaries, exams are classified as:

- (1) Normal: If both ovaries have normal morphology or simple cysts $<60\text{cm}^3$ or are not visualised but a good view of the iliac vessels is obtained, or are complex unchanged.*
- (2) Abnormal: If one or both ovaries have complex morphology or simple cysts $>60\text{cm}^3$ or ascites (vertical pool in Pouch of Douglas $>10\text{mms}$)*
- (3) Unsatisfactory: If one or both ovaries are not visualised due to a poor view. The exception is those scans where one ovary is not visualised due to a poor view but the other ovary has abnormal morphology or a simple cyst $>60\text{ cm}^3$ or $>5\text{cms}$ in diameter. In the latter case, the scan will be classified as abnormal”.*

Ultrasound Group (U) – screening strategy (pages 24,25)

“The ultrasound strategy uses transvaginal ultrasound (TVS) as a first line test with repeat screening after 6-8 weeks if an abnormality is detected on initial testing.

Level I screen: Women randomised to the U group will have an annual TVS at their TC. Depending on the results of the scan, there will be three possible courses of action (Figure 3):

- (1) *Normal scan: Routine screening with an annual TVS (Level I) on the next anniversary (one year) of the randomisation date.*
- (2) *Unsatisfactory scan: A repeat Level I scan in 12 weeks.*
- (3) *Abnormal scan: Level II scan in 6-8 weeks. Earlier scans will be arranged where there is a high index of suspicion*

Level II screen: Depending on the results of the Level II screen, there will again be three possible courses of action:

- (1) *Normal Scan: Routine screening with an annual TVS (Level I) on the next anniversary (one year) of the randomisation date.*
- (2) *Unsatisfactory scan: Repeat Level II screen in 6 weeks with earlier screens arranged where there is a high index of suspicion.*
- (3) *Abnormal scan: Referral for clinical assessment.*

Women who have repeat Level II screens will be triaged based on their findings to annual screening or clinical assessment”.

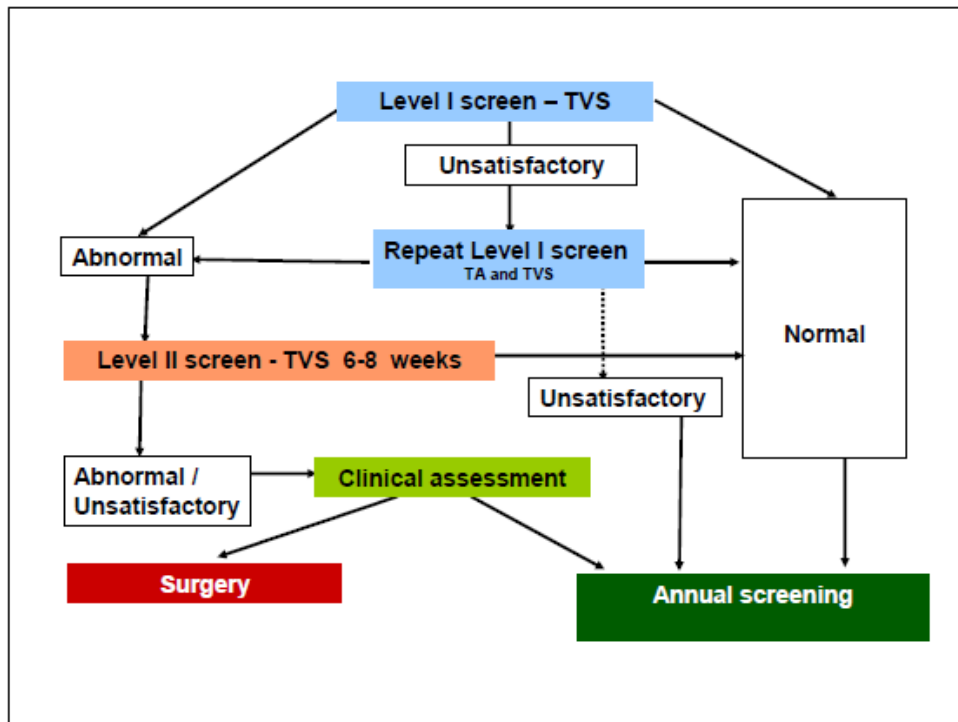


Figure 3: UKCTOCS Ultrasound Algorithm

Figure B-1: Ultrasound Algorithm (Figure 3 in UKCTOCS Protocol)

Appendix C: UKCTOCS Scanning Report Form

Appendix O - Ultrasound form
 United Kingdom Collaborative Trial of Ovarian Cancer Screening
UKCTOCS

PELVIC ULTRASOUND SCAN RESULTS

First Name _____ Surname _____ Volunteer Ref No _____
 Date ____/____/____ Ultrasoundographer _____

Please tick the relevant box

Mode of scan TRANSABDOMINAL TRANSVAGINAL BOTH Latex allergy YES NO
 Period in last year YES NO Date of last period (if within the last year) ____/____/____
 Type of HRT if used ESTROGEN PROGESTOGEN COMBINED CYCLICAL (CONVENTIONAL)
 COMBINED CONTINUOUS (NO BLEED) TIBOLONE OTHER
 Hysterectomy Previous oophorectomy NONE LEFT OOPHORECTOMY RIGHT OOPHORECTOMY

DETAILS OF OVARIAN SCAN

	LEFT OVARY / ADNEXA	RIGHT OVARY / ADNEXA
VISUALISATION MUST BE COMPLETED	<input type="checkbox"/> SEEN <input type="checkbox"/> NOT SEEN / GOOD VIEW <input type="checkbox"/> NOT SEEN / POOR VIEW <input type="checkbox"/> NOT SEEN / PREVIOUS OOPHORECTOMY	<input type="checkbox"/> SEEN <input type="checkbox"/> NOT SEEN / GOOD VIEW <input type="checkbox"/> NOT SEEN / POOR VIEW <input type="checkbox"/> NOT SEEN / PREVIOUS OOPHORECTOMY
If ovary not seen, reason	<input type="checkbox"/> BOWEL <input type="checkbox"/> FIBROIDS <input type="checkbox"/> PELVIC VARICOSITIES <input type="checkbox"/> OTHER	<input type="checkbox"/> BOWEL <input type="checkbox"/> FIBROIDS <input type="checkbox"/> PELVIC VARICOSITIES <input type="checkbox"/> OTHER
OVARIAN DIMENSIONS	mm mm mm	mm mm mm
Morphology MUST BE COMPLETED IF OVARY SEEN (If COMPLEX please write description of findings in abnormalities notes box & complete other side of this form. Also enter reference number at top of page)	<input type="checkbox"/> NORMAL <input type="checkbox"/> SIMPLE CYST <input type="checkbox"/> COMPLEX MORPHOLOGY	<input type="checkbox"/> NORMAL <input type="checkbox"/> SIMPLE CYST <input type="checkbox"/> COMPLEX MORPHOLOGY
	If midline mass, please enter under left or right adnexa and describe below If longstanding UNCHANGED complex morphology which has been previously investigated on UKCTOCS screening and is being managed conservatively, please fax form to Susan Davies on 0207 380 6929 for data entry	
Number of Cysts (more than 1 cyst = complex morphology)		
Ovary mobile?	<input type="checkbox"/> YES <input type="checkbox"/> NO	<input type="checkbox"/> YES <input type="checkbox"/> NO

Max double endometrial thickness _____ mms Fluid POD or ascites (max vertical diameter) _____ mms

Details of abnormalities: Referred in view of incidental findings YES NO

Type of Image Record NONE DISK PHOTO ONLY BOTH DISK AND PHOTO

*Result Classification /Recommended Action **Must Be Completed At Time Of Scan**

NORMAL hence ROUTINE SCREENING UNSATISFACTORY hence REPEAT LEVEL 1 SCAN
 ABNORMAL therefore LEVEL 2 / SURGERY (If any other option required, contact Susan Davies on 0207 380 6913)

• As defined by protocol.

Entered _____ Signature _____ Date _____
 Checked _____ Signature _____ Date _____

October 2005











Figure C-1: Front Page of UKCTOCS Scanning Report

United Kingdom Collaborative Trial of Ovarian Cancer Screening

UKCTOCS

Signature _____ Date _____

DETAILS OF ANY OVARIAN / ADNEXAL LESION DETECTED VOLUNTEER REF (Please enter) _____

	LEFT OVARY/ADNEXA			RIGHT OVARY/ADNEXA		
Cyst dimensions	mm	mm	mm	mm	mm	mm
Cyst wall thickness	mm			mm		
Cyst wall structure	<input type="checkbox"/> SMOOTH	<input type="checkbox"/> IRREGULAR		<input type="checkbox"/> SMOOTH	<input type="checkbox"/> IRREGULAR	
Fluid in cyst	<input type="checkbox"/> ANECHOIC	<input type="checkbox"/> RANDOM ECHOGENICITY		<input type="checkbox"/> ANECHOIC	<input type="checkbox"/> RANDOM ECHOGENICITY	
	<input type="checkbox"/> UNIFORM ECHOGENICITY			<input type="checkbox"/> UNIFORM ECHOGENICITY		
Cyst structures	<input type="checkbox"/> SEPTAE	<input type="checkbox"/> PAPILLATIONS		<input type="checkbox"/> SEPTAE	<input type="checkbox"/> PAPILLATIONS	
Maximum septa thickness	mm			mm		
Size of largest papillation	mm			mm		
Solid areas	<input type="checkbox"/> Yes <input type="checkbox"/> No			<input type="checkbox"/> Yes <input type="checkbox"/> No		
Overall impression of lesion <small>(Classification using International Ovarian Tumour Analysis criteria)</small>	<input type="checkbox"/> 	<input type="checkbox"/> Unilocular cvst		<input type="checkbox"/> 	<input type="checkbox"/> Unilocular cvst	
	<input type="checkbox"/> 	<input type="checkbox"/> Unilocular solid		<input type="checkbox"/> 	<input type="checkbox"/> Unilocular solid	
	<input type="checkbox"/> 	<input type="checkbox"/> Multilocular cvst		<input type="checkbox"/> 	<input type="checkbox"/> Multilocular cvst	
	<input type="checkbox"/> 	<input type="checkbox"/> Multilocular solid		<input type="checkbox"/> 	<input type="checkbox"/> Multilocular solid	
	<input type="checkbox"/> 	<input type="checkbox"/> Solid		<input type="checkbox"/> 	<input type="checkbox"/> Solid	

DOPPLER STUDY OF ABNORMAL AREA

Doppler performed	<input type="checkbox"/> YES <input type="checkbox"/> NO
Presence of colour signal	<input type="checkbox"/> YES <input type="checkbox"/> NO
Location of colour signal	<input type="checkbox"/> SEPTAE <input type="checkbox"/> WALL <input type="checkbox"/> SOLID AREA <input type="checkbox"/> PAPILLATIONS <input type="checkbox"/> OTHER
Lowest RI measured	_____
Lowest PI measured	_____
Peak systolic velocity	_____

Findings suggestive of DERMOID CYST ENDOMETRIOTIC CYST

Were scan images reviewed at CC? YES NO

Person at CC with whom results discussed _____

Suggested management AS RECOMMENDED OTHER (if other, please enter details in notes)

NOTES

October 2005

Figure C-2: Rear Page of UKCTOCS Scanning Report

Appendix D – UKCTOCS TVS Scan Classification Algorithm

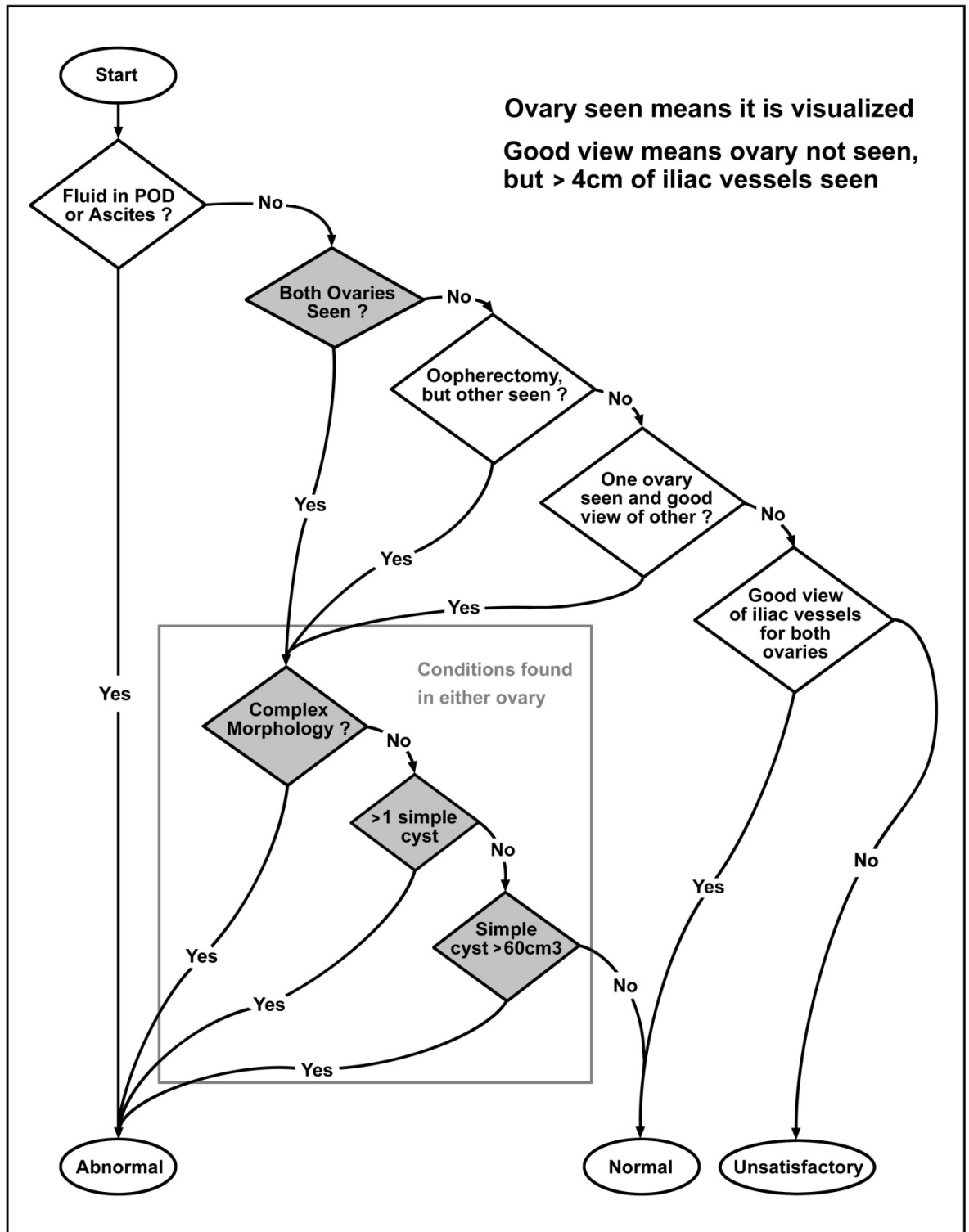


Figure D-1: UKCTOCS TVS Scan Classification Algorithm (flow diagram)

Appendix E: Ultrasound Quality Assurance in UKCTOCS

The Ultrasound Quality Assurance in UKCTOCS document was created in 2006 and is copied below verbatim with kind permission of the copyright holder, University College London.

“

Ultrasound quality assurance in UKCTOCS

Background

All screening trials need to develop quality assurance methods for the tests being evaluated if future large-scale clinical implementation is envisaged. In UKCTOCS, one of the screening tests being evaluated is ultrasound (USS). 50,000 women undergo annual transvaginal scanning at 13 UK centres. It is known that ultrasound has a significant subjective element. There is considerable inter-observer variability and accuracy of interpretation is dependent on operator experience. In addition, at the start of the trial in 2000, there were no internationally agreed definitions of various USS features. It is therefore essential to develop, implement, audit and refine quality assurance methods for ultrasound in the course of the UKCTOCS trial.

Ultrasound quality in a multicentre research trial is dependent on proper infrastructure, a well defined protocol, review and audit.

A. INFRASTRUCTURE

All centres have a dedicated ultrasound room. All scans on the trial are done on a dedicated Kretz SA9900 digital ultrasound scanner. The annual servicing contract includes one to two inspection visits and up to 8 emergency or breakdown visits including all the parts inclusive of the ultrasound probes. As part of the equipment set up, similar settings were applied to the ultrasound machine at each centre to ensure that all the images are of similar quality/clarity.

B. STAFF

There are two level of staff scanning on the trial.

Level I ultrasonographers – *They perform annual Level I scans and repeat scans following an unsatisfactory Level I scan. They are certified ultrasonographers, or trained midwives or doctors with experience in gynaecological scanning especially transvaginal scanning. Where possible the staff must be part of a NHS Radiology department or involved in scanning NHS or private gynaecological patients in addition to working on the UKCTOCS trial.*

Level II ultrasonographers – *They perform Level II scans when an abnormality is detected on the annual scan in the USS arm of the trial or when a scan is indicated in the Multimodal arm of the trial. They are experienced gynaecologists or radiologists or senior ultrasonographers (usually at superintendent grade in the NHS) with particular experience in transvaginal ultrasonography. In some centres Level II ultrasonographers perform Level I scans as well. In some centres a third opinion is sought in cases where there are equivocal findings from a senior consultant in the NHS with expertise in imaging gynaecological malignancies.*

C. TRAINING

Initial training of Level I ultrasonographers was provided on site by the senior ultrasonographers/gynaecologists (Level II scanners) at the regional centre. Each centre was then visited by the senior study investigator with ultrasound expertise who assessed the ultrasound set up, scanned with as many of the Level I scanners as possible and provided further on site training.. After all trial centres were set up, a meeting of ultrasonographers working on the trial was then arranged at the coordinating centre in London a so that various issues related to scanning on the trial and the protocol could be discussed in detail.

Ongoing trial updates and training

Regular newsletters are sent from the coordinating centre to all the staff working on the trial. Specific leaflets are sent to the ultrasonographers to update them with regard to scan definitions, classification of ovarian morphology, changes in the UKCTOCS scan form and importance of adherence to protocol. Sonographers are supervised by the senior ultrasonographer/ gynaecologist at the regional centre.

When there are any concerns based on audit, one of the study investigators visited the centre to provide further on site training.

Regular annual meetings of the lead researchers of the centres are arranged. Specific issues include to discussion of individual cases from the trial, experience of each of the centres on the issue of ultrasound, review of ultrasound protocol including definition, classification and updating of management guidelines in line with new publications in the literature.

Plans from 2006 onwards

- 1. Six monthly updates to individual ultrasonographers which will include personal audit data, details of screen negative cancers with comments for the expert reviewers on the USS subcommittee*
- 2. Regular emails to those scanning on the trial to alert them to publications of relevance to UKCTOCS*
- 3. Annual meeting of ultrasonographers working on the trial to discuss issues specific to ovarian scanning in postmenopausal women, audit findings and to review scans of missed cancers.*
- 4. Annual centre visits by USS subcommittee experts to assess on-site scanning and address training issues of individual ultrasonographers*
- 5. Setting up of a web based training programme*
- 6. Certification and annual revalidation of ultrasonographers working on the trial using a web based approach.*

D. PROTOCOL

Scanning protocol

There is a detailed scanning protocol which is provided for each sonographer. It includes clear definitions of USS features (based on IOTA classification) and ensures that any women detected to have an ovarian abnormality (all abnormal morphology except a <60cc single simple cyst) is rescanned by a Level II ultrasonographer.

Data collection

A specific ultrasound data form developed for the trial is used. The form has been audited and wording modified to ensure that there is no misunderstanding with regard to completion of the various fields in the form.

The forms are completed at the time of the scan and are entered by the regional centre via the web browser onto the central database at the coordinating centre (CC). To ensure that the data from the ultrasound scan forms is entered correctly, on site training is provided to the ultrasonographers and the research nurses. Any queries with regard to scan data entry are dealt by the senior research nurse at the coordinating centre. All problem scans are faxed over to her for data entry. Most scan data is entered within 2 days of the scan being performed. The timelines are monitored annually by DMEC.

Regional centre staff after entering the scan data, click on a 'calculate' button which then automatically classifies the scan result. Thus they are able to check that the protocol driven result classification agrees with that assigned by the ultrasonographers.

Classification of results and ensuing management

Each night, the central data management system (DMS) triages volunteers according to their scan results into protocol based management options. As all clinics in the regional centres are set up on the central DMS, appropriate appointments are made to Level I clinics if the results are unsatisfactory, to Level II clinics if an abnormality is suspected on Level I scanning and to Level I appointments in a year if results are normal. All letters advising volunteers of their result and appointment details for further scans are sent from the CC at UCL.

Deviation from protocol

Only designated CC staff can take women off protocol. The regional centre team on advice from the lead consultant or designated clinician advice CC staff if volunteers are to be suspended from the routine protocol. Such volunteers are flagged as being off protocol due to 'clinical decision'.

Second opinion

In case where there are unusual or equivocal features on ultrasound images, the regional centre request a 2nd opinion on the ultrasound features. Ultrasound images are reviewed centrally by Prof Stuart Campbell or locally in certain centres by the senior consultant radiologist at the NHS Trust.

Archiving and review of images

A unique feature of the trial is central archiving as far as possible of ultrasound images for future review and audit. There is a standard protocol on the manner in which images should be labelled and the number of views of normal and abnormal ultrasound images which should be stored. The ultrasound machine allows for local archiving. Weekly, the images are transferred onto magneto-optical discs and sent to the CC where the images are uploaded onto a central archive.

AUDIT

There is no standard format for assessing quality of ovarian ultrasound in a multicentre trial. The protocol proposed is continuously developed and refined within the trial with the objective of developing effective strategies to assess, audit and improve ultrasound quality. It is based on assessment of normal scans of women in order to explore differences between centres and operators.

The initial audits involved assessment of

- 1. Effect of various parameters such as age, HRT use, previous hysterectomy on visualisation and assessment of ovaries*
- 2. Visualisation rates of normal postmenopausal ovaries and volume of ovaries described as normal both overall, for individual centres and individual operators.*
- 3. Impact of increasing experience on working in the trial on the above parameters.*

Only annual Level I transvaginal scans were analysed, both for centres and also for individuals who had performed over 100 scans on the trial. Similarly, the ovarian volume where ovarian morphology is described as 'normal' is assessed for between-

centre variation, as well as extreme values. Any systematic variation found across centres in the women's age, BMI or incidence of hysterectomy is taken into account. Data is analysed to assess level of consistency across centres and between individual ultrasonographers. The audit is discussed at the annual ultrasonographer's meeting and reports are sent to all centre leads for dissemination to the ultrasound team. Members of the ultrasound subcommittee will visit centres to scan with specific individuals.

This was done once every two years during recruitment and is planned twice a year in the screening phase of the trial.

DMEC review of ultrasound data

Annually the ultrasound data is analysed in detail and reported to DMEC. The rates of surgery, apparent sensitivity, positive predictive value and false negative rates are examined by the independent committee for any deviation from estimates in the grant submission. Comparison is also made to any new data published from other trials especially the PLCO in the USA.

”

Appendix F: Enhanced Anonymization of TVS Images

UKCTOCS recorded the name of the volunteer in both the DICOM image and header at the time of scanning. Therefore these images had to be handled as personal medical data using the procedures prescribed in the UCL Data Protection Act (DPA) registration documents, reference Z6364106. The handling procedures required by the DPA would have severely limited the research and training potential of the images, so the trial management team required all DICOM images to be pseudo-anonymized before storage in the Ultrasound Record Archive (URA) at the coordinating centre. This work converted the images from personal medical to anonymous research data by removing the volunteer name and also ensuring the association between volunteer reference and volunteer name was not made available to researchers or the public. Other data stored in the DICOM header that may have allowed identification of the person to whom the image related, such as date of birth, were also removed.

In order to provide more secure pseudo-anonymization it is proposed that the DICOM reference^a, volunteer reference, scan date and trial centre name should all be removed in both the header and the image of the DICOM file, so that it could be identified only in terms of its image identifier. This identifier might take the form of a GUID recorded in a specific header field as well as in the image itself; see Figure F-1.

GUID Identifiers

A Globally Unique Identifier (GUID) is an implementation of the Universally Unique Identifier (UUID) standard [Ref]. It is a 128-bit value represented as a group of 8 hexadecimal digits, followed by three groups of 4 hexadecimal digits each, followed by one group of 12 hexadecimal digits; e.g. 70FBC4FF-0BA5-493C-A3D3-B61F9F656839. Typically, a GUID is generated from 122 random bits giving a numeric range of 2^{122} or 5.3×10^{36} which creates a very low probability of the same number being generated twice. Accordingly it is generally assumed to be unique.

^a The DICOM reference is a unique number, but includes the date and time of the scan as well as the ultrasound machine identifier. This information could facilitate identification of the person who was scanned.

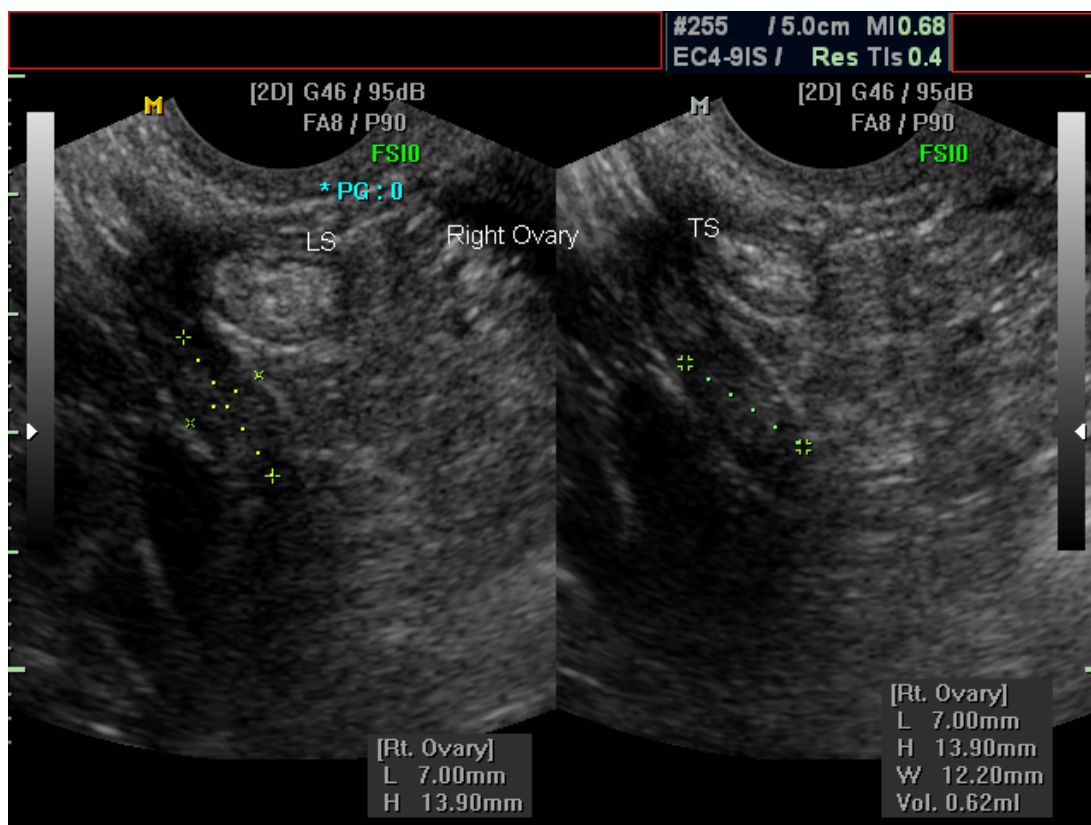


Figure F-1: TVS static image with personal identifiable information removed

Appendix G: Comparison of Traditional and Cloud Datacentres

It is estimated that a total of \$240 billion will have been invested globally in Cloud datacentres facilities by 2017²⁴⁰. Although such investment accounts for only 4% of all IT spending in 2015, Cloud computing is experiencing rapid growth in comparison with other parts of the industry, which are stagnant or declining.

The growth of Cloud Datacentres is driven by the potential cost reduction and improved quality that they should be able to deliver. Accordingly, it is anticipated that over the next ten years there will be a steep decline in the use of a traditional datacentre as they are replaced by the kind of server farm illustrated in Figure G-1.

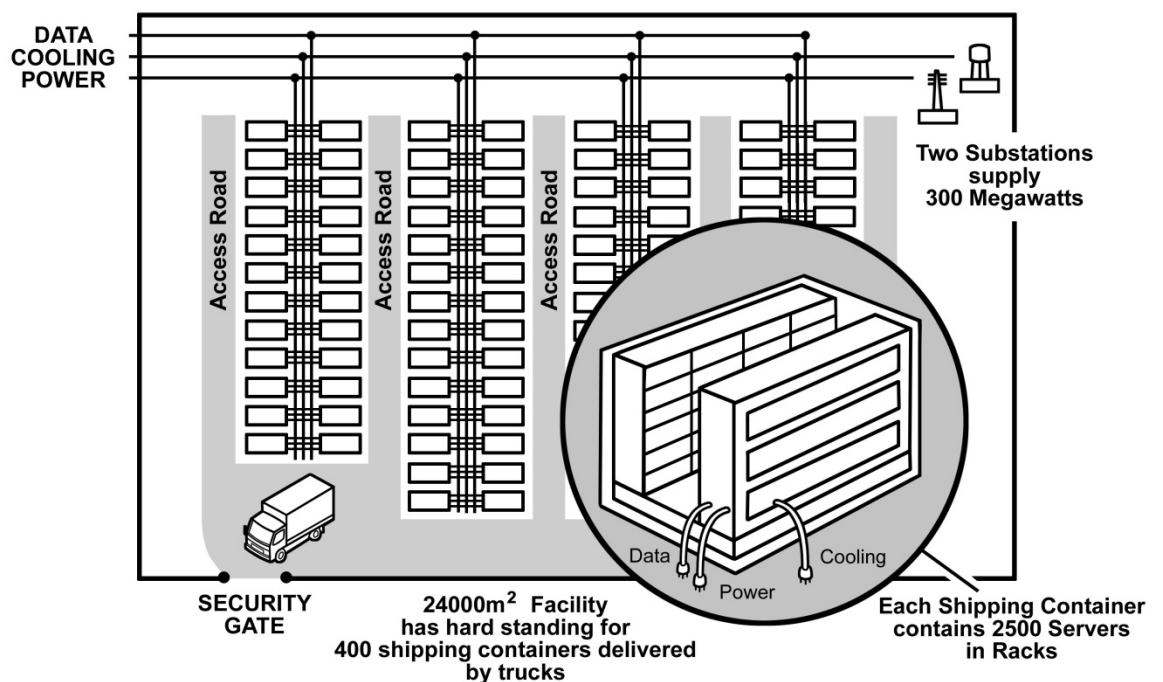


Figure G-1: Cloud Datacentre

The differences between the traditional and cloud datacentres are characterised in terms of competitive advantage in the following key areas:

1. **Size:** Traditional datacentres have racks of servers in a large air-conditioned room managed by a team of local technicians. By comparison the Cloud datacentre is much bigger, consisting perhaps of 10,000 m² of hard standing

with provision of power, network connections and cooling for hundreds of sealed shipping containers each containing hundreds of servers; see Figure G-1. These servers are remotely managed so that in the event of the failure of one server, the services can be transferred to another server in another container. When the number of failed servers in a container reaches a threshold, a truck arrives with a new container which is quickly connected to the power, network and cooling services, while the old container is returned for refurbishment.

2. **Location:** Traditional datacentres had to be built in large cities because of the need to have a ready supply of skilled technicians to operate them and the high cost of data communication mandated that they had to be close to where people used their services. The introduction of high speed and low cost data communication to large areas of the globe allows Cloud datacentres to be located in places that have cheap electricity, low labour costs, and inexpensive land. Skilled technicians are no longer required to be co-located in the Cloud datacentre so opportunities for out-sourcing can also be exploited. However, regulatory requirements (such as the need for data to be maintained within a particular geographic area), political stability and security still need to be considered.
3. **Commoditization of services:** Whilst the virtualisation of servers^a is common-place in most traditional datacentres, a wide variety of different services are run often on a number of different operating systems. For example the URA will need to be operated in the UCL datacentre on a virtual machine running Windows Server 2003 and SQL Server 8.0, whereas the TMS will need Windows Server 2008 and SQL Server 10.5. This variety dramatically increases maintenance costs as operators require skills in a wide range of operating systems and applications in order to apply the necessary updates, change configurations, and so forth. In contrast a Cloud datacentre

^a Virtualisation allows a single physical server to run a number of virtual machines (VM) each emulating a particular server platform, complete with its own operating system, hard disk and memory allocation. These VMs can be quickly started, suspended (entire state saved to disk), restarted, or terminated.

cuts costs by supporting only a small number of generic operating systems and applications which provide their user community^b services, categorized as:

- a. **Infrastructure as a Service (IaaS):** a control panel application allows users to setup, start, and stop their own virtual machines (VM). The users are usually responsible for supplying and maintaining the operating system and applications running on their VMs. The datacentre is responsible for supplying and maintaining the hardware and its infrastructure as well as the VM provision and control panel; resource pooling. Typically, the datacentre will also provide data backup and recovery, including that of the VMs themselves.
- b. **Platform as a Service (PaaS):** extends the IaaS approach by providing VMs with a standard stack comprising operating system, webserver, database and programming language; e.g. Linux, Apache, MySQL and PHP (LAMP). The software application is then developed and run by users using this stack, which is supplied and maintained by the datacentre.
- c. **Software as a Service (SaaS):** goes a step further than PaaS by providing end-users with an application supplied and maintained by an independent software vendor (ISV), but hosted by the Cloud datacentre. SaaS from the perspective of an end-user is usually no different to any other thin-client solution involving access to a website from their browser. However, SaaS normally gives the ISV an ability to bill end-users according to their resource usage as well as a facility to vary the resources provided to individual end-users, satisfying fluctuations in their demand; see elasticity. Therefore SaaS allows services to be provided more efficiently and priced on actual usage with a guaranteed level of service.

^b The user community includes the people creating applications (developers), people responsible for supporting such applications (technicians), and people who use the applications to do their work (end-users).

4. **Elasticity:** A traditional datacentre will allocate resources to users as stipulated up-front by a service level agreement (SLA). For example, users may rent a physical server for a year during which time the datacentre will supply it with electricity, cooling and network connections as well as guaranteeing to fix any hardware problems within a given timescale. Alternatively, users may elect to share a physical server with other users and share its resource (co-hosting), or just have use of a virtual machine running on a server alongside other VMs with resources allocated according to their agreement. Other options may also be included in the agreement such as the supply and maintenance of a standard operating system or application stack and the provision of network and machine load balancing when more than one server is rented. Cloud datacentres are associated with a more elastic approach to renting computing services which is not fixed up-front in the SLA. Virtual machines will be started to provide extra resources when required and stopped when the demand reduces. Data storage is essentially limitless as it can be put in huge databases with high levels of reliability and performance as well as partitioned and encrypted to ensure complete security. Data communication between different VMs, the databases and external users will also be scaled to meet demand. This flexibility in the provision of resources may be set up to happen automatically so allowing response to peaks and troughs or demand within minutes, with users only paying for what they consume.
5. **Reliability:** A traditional datacentre is vulnerable to the single points of failure which are inevitable on a single site. For example: a digger cutting through the main network cables, a terrorist incident stopping staff arriving for work, a legal injunction closing down the site. Cloud datacentres are inherently more reliable as they usually have multiple sites located in different countries with the ability to re-provision servers and re-route network traffic in the case of catastrophic failure at one site.
6. **Development, Testing and Deployment:** In a traditional datacentre, software is usually developed and mainly tested in a local network and then deployed to the datacentre for final testing and production use. This approach may result in a longer development cycle as the deployment phase can take weeks and some problems only become evident at the end of the process. A

Cloud datacentre will typically automate much of the setup and deployment of software, facilitating the rapid provision of execution environments and allowing the implementation of agile techniques such as continuous integration. This approach can dramatically reduce the time and cost of implementing and maintaining a software system whilst also providing better quality through improved testing.

In the case of a new computer system developed to exploit the benefits of a Cloud datacentre there is significant opportunity for cost savings with reduced development times and higher levels of quality, security, reliability and accessibility than can be achieved in a traditional datacentre or by employing servers in a departmental data room. However, there are a number of barriers to the use of Cloud datacentres which need to be considered, such as:

1. **Compliance with regulatory and ethical requirements:** the storage of data in a Cloud datacentre located outside the UK would require careful consideration, particularly in respect of statutory legislation and ethical requirements.
2. **Support for non-standard operating systems and applications:** the advantages of a Cloud datacentre are diminished if it becomes necessary to host a system as a IaaS with the installation and maintenance of the operating system and applications undertaken by trial staff.
3. **Integration with private networks:** it may not be possible to transfer data into the Cloud using a private network. However, private networks like NHS.NET have been integrated with services provided by Government-as-a-Cloud (G-Cloud) datacentres, including services concerned with the storage of personal medical data.
4. **High Performance Computing (HPC):** HPC is usually concerned with supporting one user using a single application to do a huge amount of processing on equally huge datasets. It is often more efficient to perform HPC processing by spreading the work across clusters of processor cores rather than different VMs (see elasticity). Therefore cloud datacentre architecture may not be suitable for HPC work. However, some Cloud vendors can support HPC work by providing machines with clustering

capability within their datacentres. This would allow VMs to be used for the systems needed to administer the trial and clusters to be used for any specialist HPC processing of the collected images.

5. **Large data transfer:** Cloud datacentres usually charge for data transfer in and out of the datacentre, but do not charge for data transfer within the datacentre. This charging model benefits a trial like UKCTOCS in which <1.5 million relatively small images (<1MB) need to be transferred into the datacentre and only a small proportion will subsequently be transferred out. However, this sort of charging model might not work for a trial involving a higher number of large images than need to be frequently moved in and out of the datacentre.
6. **Government access to data:** Cloud datacentres centralise data storage, which facilitates monitoring by government agencies. This may not be an issue for state sponsored medical research, but can have implications for the storage of commercially sensitive data particularly in the case of datacentres located in a region which is an economic competitor of the region from which the data owner originates.

It is possible that the rise of the Cloud datacentre heralds a further significant change in the way computing is provided. In the much same way that the microprocessor brought computers into every home, the cloud might usher in the measurement and control of almost every aspect of modern life; the internet of things.

Appendix H: DICOM File Structure

The images contained in the URA are stored in DICOM format so their analysis involves reading the data in this format and possibly converting it into another format more suited to processing by the workflow. Therefore it is necessary to obtain an understanding of the format of the data in a DICOM image.

DICOM Standard

The Digital Imaging and Communications in Medicine (DICOM) format has become the leading format for the communication and storage of medical images. It is the result collaboration between a group of people who use medical images and representatives of the main equipment manufacturers, respectively the American College of Radiology and the National Electrical Manufactures Association (NEMA). The DICOM format is governed by an international standards committee. The current version is PS3.1 2016b and it is publicly available without fee from the NEMA website²⁴¹.

At the highest level the DICOM standard defines the fundamental services and objects it supports (parts 4 and 3, respectively). Services include image storage, query and retrieval as well as patient management and procedure steps. Objects include metadata (patient information, machine setting, etc.), image types (CT, MRI, Ultrasound, etc.) and similar information which is defined in terms of specific type definitions (Part 5) and individual attributes specified by a data dictionary (Part 6). The standard also details its role in both the communication and storage of all common types of medical images and serves to facilitate interoperability between the equipment of different manufacturers.

In terms of communication DICOM describes only the session, presentation and application layers of the OSI Seven Layer Model²⁴² as it relies on TCP/IP²⁴³ and Ethernet²⁴⁴ for the transport, network, data link and physical layers. In terms of storage, part 11 of the standard defines the specific application profiles for image modalities (cardiology, ultrasound, etc.) with parts 10 and 11 respectively defining the file format and supported types of physical media. The standard also defines data

structures, data types and codes which are common to the roles of DICOM in both the communication and storage of images.

The DICOM standard is necessarily extensive and detailed, as it must encompass the huge variation in the application of imaging that exists in modern medical practice. However, only a small subset of the standard is important for the purposes of processing UKCTOCS images stored in the URA. These images are each stored as a Binary Large Object (BLOB) field in a record of a relational database table. The BLOB contains serialised bytes corresponding to a DICOM file. The format of the file is fully specified by the DICOM standard, but is essentially composed of a header data structure followed by an image data structure.

Header Data Structure

The header data structure starts with a few bytes that serve to define the file as containing data in the DICOM format. It then contains information about the size and content of the header. This allows calculation of the start location of the image data structure in terms of an offset (number of bytes) from the start of the file. The next part of the header defines the image modality which dictates the associated metadata. Objects in the metadata are composed of a type and an element count. Some metadata types have just one element, for example the patient's surname. Other metadata types have multiple elements like a list of the patient's other names. The elements themselves are also specified to have particular data type or structure so that, for example, numeric values can be expressed as four byte integers and strings as a count followed a sequence of encoded characters. In this way the objects in the metadata can be located by traversing the data structures starting from one object's offset and then adding the length of its data structure to find the offset of the next object.

The development of the URA system (chapter 4) revealed that the DICOM files created by UKCTOCS had no header information that would be helpful during any subsequent analysis apart from identification of the volunteer and scan date; see Figure 7-A-1. In particular, they lacked information about their use in measuring ovaries as well as the location of any calliper marks in the image and the distance

between them. Therefore there was no requirement to read the DICOM header beyond using it to obtain the offset of the image data structure.

Tag	Name	Action	Value
(0008,0008)	Image Type	Keep	ORIGINAL\PRIMARY\
(0008,0016)	SOP Class UID	Keep	1.2.826.99999.1.4.1.1.6.1
(0008,0018)	SOP Instance UID	Keep	1.2.826.999999.1.9999.1.1.8.0
(0008,0020)	Study Date	Keep	
(0008,0021)	Series Date	Keep	
(0008,0023)	Content Date	Keep	
(0008,0030)	Study Time	Keep	
(0008,0031)	Series Time	Keep	
(0008,0033)	Content Time	Keep	
(0008,0050)	Accession Number	Keep	
(0008,0060)	Modality	Keep	US
(0008,0070)	Manufacturer	Keep	MEDISON
(0008,0080)	Institution Name	Keep	UKCTOCS
(0008,0090)	Referring Physician's Name	Keep	
(0008,1010)	Station Name	Keep	
(0008,1030)	Study Description	Keep	
(0008,1070)	Operators' Name	Keep	
(0008,1090)	Manufacturer's Model Name	Keep	ACCUVIX
(0010,0010)	Patient's Name	Keep	xxx
(0010,0020)	Patient ID	Keep	
(0010,0030)	Patient's Birth Date	Keep	
(0010,0040)	Patient's Sex	Keep	0
(0018,1000)	Device Serial Number	Keep	
(0020,0000)	Study Instance UID	Keep	1.2.826.99999.1.9999.1.1.8.0
(0020,000E)	Series Instance UID	Keep	1.2.826.99999.1.9999.1.1.8.0.1
(0020,0010)	Study ID	Keep	
(0020,0011)	Series Number	Keep	1
(0020,0013)	Instance Number	Keep	8
(0020,0020)	Patient Orientation	Keep	
(0028,0002)	Samples per Pixel	Keep	3
(0028,0004)	Photometric Interpretation	Keep	RGB
(0028,0006)	Planar Configuration	Keep	0
(0028,0010)	Rows	Keep	480
(0028,0011)	Columns	Keep	640
(0028,0100)	Bits Allocated	Keep	8
(0028,0101)	Bits Stored	Keep	8
(0028,0102)	High Bit	Keep	7
(0028,0103)	Pixel Representation	Keep	0
(0028,1050)	Window Center	Keep	128
(0028,1051)	Window Width	Keep	256

Figure H-1: Contents of DICOM Header from UKCTOCS Image

Image Data Structure

The static images obtained during TVS examination were 640 pixels (wide) by 480 pixels (high) and contained 8 bits per pixel giving 256 levels of echogenicity in terms of a grey-scale image. Therefore the image data structure starts with a count of the image width and height and the subsequent series of bytes form an image such that each pixel occupies one byte and each row of the image contains 640 bytes. In addition images are able to contain colour information, such as might be obtained by recording Doppler flow information. This is achieved by arranging the data as a sequence of three grey-scale images each representing one of the primary colours; red, green, blue. In this way the three images can be combined to produce a full-

colour image. However, the parts of the image that do not contain colour information have the same grey-scale values in each of the three image data structures, so a 256 level grey-scale representation is formed when the red, green and blue values are combined in a colour display.

Knowledge about the image data structure allows it to be analysed for the purposes of finding objects such as calliper marks and ovary boundaries. The ability to perform such analysis is a key requirement of the processing performed by the image workflow.

Appendix I: Processor Types Used in Ultrasound

Signal Processing

A variety of different processor types are employed at various stages in the processing of ultrasound signals as shown in Figure 2-6. The following list summarises their design and function:

- **Graphics Processor Unit (GPU)** – contain multiple processor cores, large amounts of memory and high speed buses to allow the high speed transfer and parallel processing of data. For example, NVIDIA Tesla M60 has 4,096 thread processors, 16GB RAM and a bus speed of 160 GB/s. It can perform up to 300×10^9 single precision operations per second (300 GFLOPS) and is optimised for performing kernel image processing tasks which can be programmed in C++ using the CUDA application programming interface.
- **Digital Signal Processor (DSP)** – processors optimised for computation-intensive functions implemented using matching C/C++ compilers generating the required DSP machine codes. They are commonly supplied as a complete system on a chip or board by combining DSP with other processor running a real-time operating system. For example, Texas Instruments C6678 has 8 DSP cores, 4MB internal memory and can perform 160×10^9 single precision operations per second. DSPs are particularly to applications that perform iterative operations on data, such as might be required for the implementation of a Fast Fourier Transform (FFT) – i.e. decomposing a sequence of values into components of different frequencies.
- **Application-specific Integrated Circuits (ASIC)** – an integrated circuit customised for a particular purpose, for example beamforming. Typically they are developed using a hardware description language (HDL) which describes the features to be fabricated on the chip in terms of the required photolithographic layers. ASICs can contain complete system on a chip as well as analogue components which makes them ideal for replacing circuit board implementations in front-end electronics without significant degradation of performance. They are particularly to applications that

perform transformations on a flow of data rather the type of complex processing that typically performed by a DSP.

- **Field Programmable Gate Array (FPGA)** – an integrated circuit that can be customised after manufacture using a HDL similar to that used for developing ASIC chips. It contains an array of programmable logic blocks and memory elements allowing the implementation of complex digital processing, but with performance typically inferior to that delivered by the equivalent ASIC.

Appendix J: Review of Image Processing Tools

3DSplicer (SPL, Harvard)

Origin

The Surgical Planning Laboratory²⁴⁵ (SPL) is a part of the Department of Radiology at the Brigham and Women's Hospital (BWH) which is affiliated with Harvard Medical School, USA. It undertakes research and development in image processing algorithms, software systems and medical applications. It received funding in 1997 to create a new Neuroimage Analysis Centre (NAC) on the BWH campus. The NAC is currently funded by from the U.S. National Institute of Biomedical Imaging and Bioengineering (NBIB) and is focused on the processing of diagnostic imaging data. This led to the creation of the National Alliance for Medical Image Computing²⁴⁶ (NAMIC) to provide the infrastructure and environment for the development of computational algorithms and open-source technologies. The core leadership of NAMIC is provided by BWH SPL.

Development Process

NAMIC publishes various standard image datasets (Brain, Prostate, Heart) as well as the Slicer software application and associated toolkits²⁴⁷. It also has a remit to develop software engineering methodologies to support its work though this seems to comprise of little more than suggesting the use of standard tools²⁴⁸ (CMake, CPack, CTest, CDash) and Extreme Programming methods¹³⁶. The organisation of NAMIC is summarised in Figure J-1 and contains a number of elements that are similar to the organisational environment discussed in Figure 6-1.

The NAMIC structure recognises the separation of software engineers and algorithm developers as suggested in Chapter 6, but doesn't differentiate between users who are annotating datasets using their specialist domain knowledge (clinicians) and users who are performing research. This is an important segmentation in terms of the type of tools needed to support work in these two areas and how they interoperate. For example, clinicians have the expertise to identify ovary boundaries so need a tool that allows them to annotate TVS images by creating an overlay showing the location of the boundary. In this respect Chapter 5 describes ImageCurator, a tool

that displays an image and allows the boundary to be drawn using a touch-sensitive screen and pen device. The boundary locations are recorded as a series of Cartesian coordinates in an IDF file which allow an overlay to be created without altering the source image. The annotations in the IDF file can then be used by image researchers in other tools concerned with conducting image experiments, such as ImageExpter described in Chapter 6. In this way the boundaries identified by the clinicians can be compared with the boundaries found by the image researchers as shown in Figure 5-4. The structure devised by NAMIC does not recognise this key differentiation between their users and this is reflected in the absence of specialised tools for particular user segments in the products they publish. However, NAMIC does recognise the importance of separating the work of algorithm developers from the work of application developers who build tools to support users, shown respectively as Computer Scientists and Engineers in Figure 6-1. This is depicted in Figure J-1 by algorithm developers producing software libraries (Visualisation Toolkit (VTK), Insight Toolkit (ITK), Teem Libraries (NRRD)) and application developers producing modules in order to extend the 3D Slicer core application.

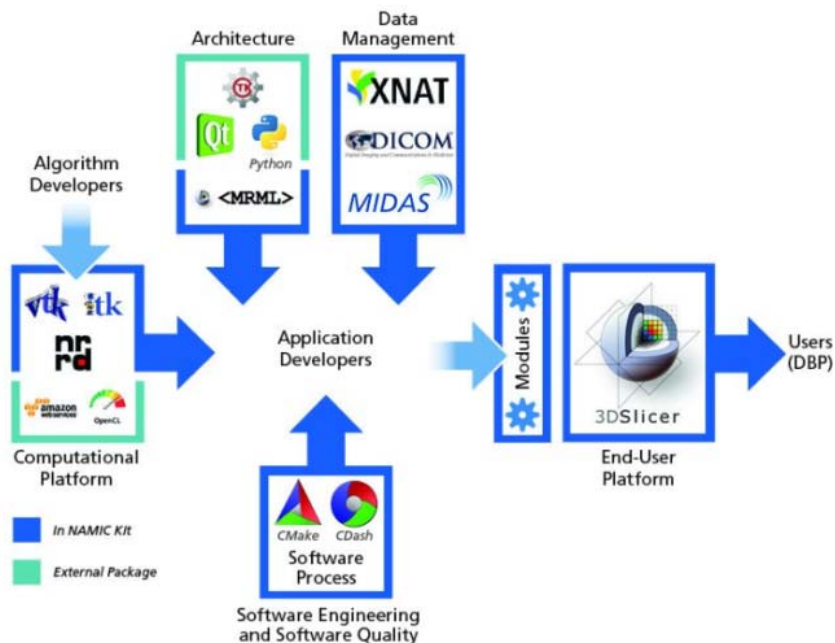


Figure J-1: Organisation of NAMIC activities © Harvard Medical School.

Tool Description

The Slicer documentation²⁴⁹ describes three types of modules:

- **Command Line Interface (CLI) modules** – complete and separate programs that can be run from the operating system command line (or by Slicer) with an interface implemented using standard command line parameters.
- **Scripted Modules** – Python scripts that can be loaded and run by Slicer and have full access to its API as well as the toolkits and libraries.
- **Loadable Modules** – C++ plugins that extend Slicer and have access to its API, toolkits and libraries.

Loadable modules have the closest correspondence to the MxPlugins described in Chapter 6. However, they lack a binary compatible interface so suffer from the problems of distributed software component development outlined in Chapter 6 which are addressed by MxPlugins. This makes it difficult for people outside NAMIC to develop Slicer modules as they need to use the same tools, follow the same build process and become closely involved with their developer community. It also means that any intellectual property must be disclosed as the source code for any module needs to be added to the Slicer code base. Consequently, it seems unlikely that NAMIC will create the sort of marketplace for image processing components as envisaged in Chapter 6. This is confirmed by the small number of contributors acknowledged on the Slicer website who are not part of SPL and the absence of any list of loadable modules developed outside SPL.

The CLI and Scripted modules are not implemented in the current version of ImageExpt, but will be considered for future development as they provide a middle ground between non-programmers who create processing workflows using the graphical DSL (Figure 6-7) and experienced C++ developers who create custom process steps by building an MxPluginForImages component.

Workflow Capabilities

The Slicer website did not provide details about implementing a workflow suitable for processing thousands of images, though given its origins in MRI and cranial imaging this must be possible. Therefore is it not possible to compare Slicer with the

image workflow tools described in Chapter 6 or determine its suitability for image transaction processing. Even so the environment proposed in Chapter 6 for research involving image analysis still has clear advantages over the approach currently taken by NAMIC in a number of key areas, as outlined above. However, the work of NAMIC has the advantage of being a mature platform that is well funded and has an established community of both developers and users. It remains to be seen as to whether the ideas presented in Chapter 6 will prove attractive enough to attract similar levels of support.

MeVis (CeVis, Bremen)

Origin

MeVisLab²⁵⁰ was originally created by Fraunhofer MEVIS²⁵¹ which is now a commercial spin-out from the Centre for Complex Systems and Visualisation (CeVis) at Bremen University, Germany. MeVisLab is currently owned by MeVis Medical Solutions AG, another CeVis commercial spin-out. This arrangement allows Fraunhofer MEVIS to concentrate on science and research whilst MeVis Medical Solutions concentrates on the marketing and sales of the MeVisLab product. However, both companies are involved in the on-going development of MeVisLab.

Development Process

MeVisLab is a commercial product so does not allow third-parties to freely contribute to its code base. Therefore details of its development process are not disclosed.

Tool Description

MeVisLab is a modular framework for image processing research and development with a special focus on medical imaging. It promises fast integration and testing of new algorithms and the development of clinical application prototypes. MeVisLab provides an interface to 3D Slicer (described above), but uses the third-party application framework Qt to provide its own user interface. The Qt product is only available under a commercial licence which explains why it cannot be used by the

open-source 3D Slicer. Therefore the key differentiators between the two products are the user interface, level of support and facilities for development. In the context of Chapter 6 only the way MeVisLab can be developed for custom image processing is significant.

Development in MeVisLab²⁵² is described in terms of:

- Visual level – this is a graphical interfaces for defining image processing steps, much like the DSL described for ImageExpiter – see Figure 6-7.
- Scripting level – allows the creation of Python modules so may just be a more convenient user interface for the feature implemented by 3D Slicer.
- C++ Level – claims to allow new algorithms to be integrated ‘using the modular, platform-independent C++ class library’

MeVisLab provides scant documentation for the C++ level development in comparison with the 3D Slicer documentation. However, it does appear that MeVisLab does allow the creation of modules which are a binary compatible. This seems to be achieved by creating a ML Module which has a definable collection of inputs and outputs, each with defined types and uses²⁵³. The documentation describes only how the ML module Wizard is used to specify these definitions and gives no indication of how they are implemented. However, upon the assumption that a COM like interface has been used (see Chapter 6), it should be possible to create additional libraries that can be loaded without the need to rebuild MeVisLab. This would address a key limitation of 3D Slicer in terms of encouraging a market for image processing components. However, the commercial nature of MeVisLab would act as an impediment as any market would be limited to its users and controlled by MeVis Medical Solutions AG. It seems unlikely that many academics would be attracted to develop their algorithms for such a product.

Workflow Capabilities

It would be necessary to perform some exploratory development in order to make a proper comparison of MeVisLab to the tools described in Chapter 6. However, it is encouraging to discover that MeVisLab may have taken a similar route in terms of the architecture and design of modules that allow for its extension. This would validate the approach taken in Chapter 6 and indicate the viability of providing a

competing open-source competitor. The capability of using MeVisLab to process large numbers of image transactions is unknown, but it can be presumed that the same core code as 3D Slicer would be used (if it exists). The failure of both 3D Slicer and MeVisLab to publish assessable documentation about their respective capabilities in this respect is regrettable as this is a key facility of the tools described in Chapter 6.

XNAT (NRG, Washington)

Origin

XNAT²⁵⁴ was initially developed by a team working at the Buckner Lab at Washington University. It is currently an open-source project run by a team of developers who are mostly employed by the Neuroinformatics Research Group²⁵⁵ (NRG) at Washington University.

Development Process

The source code of XNAT is held by Atlassian Bitbucket²⁵⁶ (commercial repository based on GIT²⁵⁷) which suggests that it would be possible for developers outside the NRG team to make contributions to the product. However, the developer section of the XNAT website²⁵⁸ doesn't document the team's development process, its build cycle or the mechanism for code submission. Therefore in practice it would appear that development is restricted to the NRG team, though a small group of closely associated helpers from outside this group are mentioned on the website. XNAT does promote a marketplace website²⁵⁹ for the distribution of third party tools and plugins, but only 6 tools and 24 are listed.

Tool Description

XNAT is a data archiving tool for neuroimaging projects. It supports DICOM images and delivers features for their anonymization as well as for the removal of protected health information from header data. In addition access and permission controls are provided to ensure the images remain secure. The tool also allows search and reporting on images and their associated data as well as pipeline processing, as described in the next section.

The core technologies of XNAT are Java SDK²⁶⁰, Apache Tomcat²⁶¹ and PostgreSQL²⁶² which respectively provide facilities for the building and running of application modules as well as a relational database. The developer website doesn't provide any advice for running C++ code from XNAT so it is presumed that it is not supported. However, XNAT does provide a REST API²⁶³ which would allow the development of separate applications written in C++.

Workflow Capabilities

Workflow capabilities are provided by XNAT in the form of a pipeline engine²⁶⁴ which is implemented as a standalone Java application. The process workflow is defined in an XML document which describes its steps in terms of separate XML documents defining specific executables which communicate with XNAT using parameters and standard operating system streams (STDIN, STDOUT, STDERR). The XNAT documentation does not describe the sort of transaction processing features discussed in Chapter 6.

eMedLab

Origin

Patients and medical research stand to benefit enormously from the vast amount of human health data generated from both recent and future technological advances. However, delivering such benefits depends on the development of systems that are able to meet the complex challenges of storing and processing this data in order to provide researchers with greater insights into its content and meaning. In order to address these challenges a consortium of Europe's leading biomedical scientists have obtained funding to create eMedLab. Eventually it will become a high-performance computing facility staffed by people with the skills and expertise necessary to build a world-class resource. However, the eMedLab²⁶⁵ website gives no information about its current status.

Appendix K: Data Quality

It has been recognised since the earliest days of building programmable computing devices that the right answers will not be generated if the wrong figures are provided as input²⁶⁵. This is commonly stated in terms of the phrase ‘garbage-in, garbage-out’. Therefore it is clear that any sort of meaningful processing of data is dependent on obtaining quality data. Typically the quality control of data is domain specific, but some work has been performed on attempting to identify general concepts about data quality, particularly in respect of databases. Unfortunately, a search of the literature revealed nothing that linked such general concepts to the domain of TVS image data. However, the following reviews of material about data quality in the domains of Geospatial Databases and Big Data suggest some directions for future research that might be beneficial to the type of TVS Image database described in Chapter 4.

Data Quality Parameters²⁶⁶ (H. Veregin)

The chapter reviews data quality parameters within the context of Geospatial Databases. It starts by providing a general definition of data quality and makes the point that ‘*data consumers can therefore use the same diligence in selecting a database that they might in purchasing an automobile or a pair of shoes*’. It identifies area of concern for geospatial databases including:

- Increased participation of private companies that are not required to conform to quality standard adopted by public organisations like US Geological Survey or UK Ordnance Survey
- Increased use of GIS as a decision support tool with the associated risk of litigation
- Increased reliance on secondary data sources

It also notes that responsibility for data quality has moved from the producer to the user. Indeed a central theme of the chapter is the need for database producers to specify their product in terms of ‘fitness-for-use’ and providing data quality documentation which is termed truth-in-labelling. The components of data quality for geospatial databases are categorised as spatial, temporal and thematic. It is noted that time is particularly critical to an understanding of geographical

phenomena in terms of events that appear and disappear in space and time. It is also noted that geographical phenomena are not really about space, but about theme – that is say space provides a framework for the measurement of theme without which there is only geometry. The Chapter then reviews accuracy, precision, consistency and completeness in respect of these components.

Veregin introduces definitions for accuracy, precision, consistency and completeness some of which are different from those provided in other areas of science and engineering. For example accuracy is defined as a relative measure which is the inverse of error, precision is defined as the amount of detail that can be discerned and completeness is defined as the relationship between the objects in the database and the abstract universe of all such objects. However, the definition of consistency as the absence of apparent contradictions in a database is more readily recognised by people who work outside his field. The descriptions of data quality in respect of these definitions are also very specific to geospatial databases. For example temporal accuracy is described in terms of agreement between encoded and ‘actual’ temporal coordinates and further defined as currentness or the degree to which a database is up-to-date. It is also explained that a value is current if it is correct in spite of any possible time-related changes in value.

The chapter goes on to discuss data quality standards and the following are introduced; Spatial Data Transfer Standard, Content Standards for Digital Geospatial Metadata. The work of the Federal Geographic Data Committee in promoting and disseminating such standards is also discussed. A number of other standards are also mentioned; National Transfer Format, Digital Geographic Information Exchange Standard, International Hydrographic Organisation standard. However, it is noted that such standards do not necessarily lend themselves to specific software implementations due to the absence of reference implementations. Furthermore it is observed that few commercial GIS packages offer the capability to document data quality.

Cartographic bias is discussed – the ability to produce a geospatial database presupposes a model that defines rules for simplifying real-world complexity and it is observed that despite their apparent sophistication they reflect many of the

same biases as analogue cartographic data. The chapter concludes with a discussion about GIS, Society and Data Quality. Amongst the observations made are:

- *‘GIS has led to the ascendance of a new geospatial science focused on the goal of producing ultimately truthful and objective representations of reality. This goal is seen as a by-product of the new technological means with its appeals to neo-positivism, reductionism, instrumentalist thinking, and naïve empiricism in which ‘reality’ is uncontested and objectively measurable’.*
- *‘what is contained in a database is a function not only of the nature of the external environment but also the values of the society and institution within which the database was constructed’*
- *‘given the dependence on social values it is not possible to distinguish between competing representations of the same geophysical space. Thus it has been argued that the distinction between propaganda and truth is artificial and must be dismantled, as must the arbitrary dualism between art and science’.*

The relevance of the above to data quality is not clear.

The Challenges of Data Quality and Data Quality Assessment in the Big Data Era²⁶⁷

The paper starts by giving an overview of the significant technological changes that have occurred in IT since the start of the 21st century; cloud computing, internet of things, social networking. It is observed that these technologies have resulted in a large increase in the amount and speed of data accumulation which is described as the coming of big data. The proposition is then made that in order to generate value from big data it must be accurate and high quality.

The second part of the paper is a literature review which includes work from the Total Data Quality Management Group at MIT, led by Prof Richard Wang. They define ‘data quality’ as ‘fitness for use’ and say that data quality judgement depends upon data consumers as measured by a set of ‘data quality attributes’ (or

dimensions) that each representing an single aspect or construct of data quality. Fifteen such dimensions are identified in four categories. The review also includes work by Alexander and Tate which describes six evaluation categories; authority, accuracy, objectivity, currency, coverage/intended audience and interaction/transaction features for web data. Further dimensions, evaluation categories and metrics are given in other sources. The conclusion of the review summarises studies of data quality as either involving web data or specific areas such as biology, medicine, geophysics, etc.

The third part of the paper considers the challenges of data quality in the big data era. It starts by stating the characteristics of big data as Volume, Velocity, Variety and Value. The challenges of data quality in these regards are given as:

- Diverse sources bringing abundant data types and complex data structures which are difficult to integrate
- Need to judge data quality within a time frame the corresponds to the volume being received
- Processing data before it becomes stale
- Lack of data quality standards and research on the quality of big data

The fourth part of the paper considers the quality criteria of big data and makes the point that ‘only data that conform to the relevant user and meet business requirements can be considered qualified (or good quality) data’. The difficulty of measuring data quality is noted and a hierarchical data quality standard is proposed from the user perspective. However, more useful diagram is given that categorises 14 data quality dimensions in five groups; availability, usability, reliability, relevance and presentation quality. These dimensions are then described.

The fifth part of the paper is concerned with quality assessment process for big data. This is presented as a flow diagram which is then described in detail. It is noted that in different business environments the selection of data quality elements will differ. For example for social media data timeliness and accuracy are two important quality features. It also noted that the field of biology is an important source of big data, but lacks standard in terms of data storage and

formats so consistency cannot be regarded as a quality dimension. Further insights are provided, such as the many ways in which data might be collected (data integration, search-download, web crawlers, agent methods, carrier monitors) and need for data cleaning to detect and remove errors and inconsistencies which is achieved by manual methods, specially written programs, general checking of data fields and domain specific data scrubbing.

The paper concludes that ensuring big data quality is a major issue for industry and academia and makes the point that poor quality data will lead to low data utilization efficiency and even bring serious decision-making mistakes.

Data Quality – Wikipedia²⁶⁸

Data quality is defined as being considered high quality if ‘they are fit for their intended uses in operations, decision making and planning’. A number of other definitions are also given.

The Wikipedia traces the history of Data Quality from early mainframes used by postal services to check addresses and the National Change of Address Registry. An overview of data quality is then provided. This includes a reference to ‘Zero Defect Data’ (Hansen 1991) which adapts statistical quality control to data quality. References to a number of other frameworks for data quality are also given including one to MIT Total Data Quality Management Programme led by Prof Wang.

A general approach to data quality assurance is given as a series of steps; data profiling, data standardisation, geocoding, matching or linking, monitoring and the development of batch processes to automate the steps. Data quality control is defined as the process of controlling the usage of data with known quality measurement for an application or process; severity of inconsistency, incompleteness, accuracy, precision and missing/unknown. A number of Data Quality Checks are also proposed; completeness and precision, validity, accuracy, consistency, timeliness, reasonableness, conformity, integrity.

References

References

- 1 Jacobs, I. J, Usha Menon, Andy Ryan, Aleksandra Gentry-Maharaj, Matthew Burnell, Jatinderpal K Kalsi, Nazar N Amso, et al. 2016. 'Ovarian Cancer Screening and Mortality in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): A Randomised Controlled Trial'. *The Lancet*. Accessed January 5. doi:10.1016/S0140-6736(15)01224-6.
- 2 Sharma, A., M. Burnell, A. Gentry-Maharaj, S. Campbell, N. N. Amso, M. W. Seif, G. Fletcher, et al. 2016. 'Quality Assurance and Its Impact on Ovarian Visualization Rates in the Multicenter United Kingdom Collaborative Trial of Ovarian Cancer Screening (UKCTOCS)'. *Ultrasound in Obstetrics & Gynecology* 47 (2): 228–35. doi:10.1002/uog.14929.
- 3 Timmerman, D., P. Schwärzler, W. P. Collins, F. Claerhout, M. Coenen, F. Amant, I. Vergote, and T. H. Bourne. 1999. 'Subjective Assessment of Adnexal Masses with the Use of Ultrasonography: An Analysis of Interobserver Variability and Experience'. *Ultrasound in Obstetrics and Gynecology* 13 (1): 11–16. doi:10.1046/j.1469-0705.1999.13010011.x.
- 4 NHS England Diagnostic Imaging Dataset (DID) 2013-14 [s3.5.5]
<https://www.england.nhs.uk/statistics/statistical-work-area/diagnostic-imaging-dataset>
- 5 Cancer Research UK Website:
<http://info.cancerresearchuk.org/cancerstats/types/ovary/mortality/>
- 6 <http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer/incidence#heading-One>
- 7 Eurocare - Survival of Cancer Patients in Europe. 2015. Accessed March 8.
<http://www.eurocare.it/AboutUs/tabid/54/Default.aspx>.
- 8 <http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer#heading-Two>
- 9 Walker H. Kenneth, W.Dallas Hall, and J.Willis Hurst, *Clinical Methods: The History, Physical and Laboratory Examinations*, 3rd ed. (Butterworth-Heinemann Ltd, 1990).
- 10 Randall, Leslie M., and Bhavana Pothuri. 2016. 'The Genetic Prediction of Risk for Gynecologic Cancers'. *Gynecologic Oncology* 141 (1): 10–16. doi:10.1016/j.ygyno.2016.03.007.
- 11 Tortora Gerard J., and Bryan H. Derrickson, *Principles of Anatomy and Physiology*, 12th ed. (John Wiley & Sons, 2008).
- 12 Kurman Robert J., Brigitte M. Ronnett, and Lora Hedrick Ellenson, *Blaustein's Pathology of the Female Genital Tract*, 6th ed. (Springer, 2011).
- 13 Timmerman D., et al., "Terms, definitions and measurements to describe the sonographic features of adnexal tumours: a consensus opinion from the International Ovarian Tumour Analysis (IOTA) Group," *Ultrasound in Obstetrics & Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology* 16, no. 5 (October 2000): 500-505
- 14 Florkowski, Christopher M. 2008. 'Sensitivity, Specificity, Receiver-Operating Characteristic (ROC) Curves and Likelihood Ratios: Communicating the Performance of Diagnostic Tests'. *The Clinical Biochemist Reviews* 29 (Suppl 1): S83–87.

References

- 15 Geomini P et al., “The Accuracy of Risk Scores in Predicting Ovarian Malignancy A Systematic Review,” *OBSTETRICS AND GYNECOLOGY* 113, no. 2 (February 2009): 384-394.
- 16 Jacobs I., et al., “A Risk of Malignancy Index incorporating CA-125, Ultrasound and Menopausal Status for the Preoperative Diagnosis of Ovarian Cancer,” *British Journal of Obstetrics and Gynaecology* 97, no. 10 (October 1990): 922-929.
- 17 Timmerman D., et al., “Artificial neural network models for the preoperative discrimination between malignant and benign adnexal masses,” *Ultrasound in Obstetrics & Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology* 13, no. 1 (January 1999): 17-25.
- 18 Sokalska A. et al., “Diagnostic accuracy of transvaginal ultrasound examination for assigning a specific diagnosis to adnexal masses,” *Ultrasound in Obstetrics and Gynecology* 34, no. 4 (10, 2009): 462-470.
- 19 Brown, DL., KM Dudiak, and FC Laing, “Adnexal Masses: US Characterization and Reporting,” *RADIOLOGY* 254, no. 2 (February 2010): 342-354.
- 20 Vergote I., et al., “Prognostic importance of degree of differentiation and cyst rupture in stage I invasive epithelial ovarian carcinoma,” *LANCET* 357, no. 9251 (January 20, 2001): 176-182.
- 21 Meinhold-Heerlein, Ivo, Christina Fotopoulou, Philipp Harter, Christian Kurzeder, Alexander Mustea, Pauline Wimberger, Steffen Hauptmann, and Jalid Sehouli. 2016. ‘The New WHO Classification of Ovarian, Fallopian Tube, and Primary Peritoneal Cancer and Its Clinical Implications’. *Archives of Gynecology and Obstetrics* 293 (4): 695–700. doi:10.1007/s00404-016-4035-8.
- 22 Prat, Jaime, and for the FIGO Committee on Gynecologic Oncology. 2015. ‘Abridged Republication of FIGO’s Staging Classification for Cancer of the Ovary, Fallopian Tube, and Peritoneum’. *Cancer* 121 (19): 3452–54. doi:10.1002/cncr.29524.
- 23 Kosary C (2007) *Cancer of the Ovary*. Reis LAG et al ‘Cancer Survival Among Adults: U.S. SEER Program, 1988-2001. http://seer.cancer.gov/archive/publications/survival/seer_survival_mono_lowres.pdf.
- 24 Kurman, Robert J., and Ie-Ming Shih. 2016. ‘The Dualistic Model of Ovarian Carcinogenesis: Revisited, Revised, and Expanded’. *The American Journal of Pathology* 186 (4): 733–47. doi:10.1016/j.ajpath.2015.11.011.
- 25 Brown, Patrick O., and Chana Palmer. 2009. ‘The Preclinical Natural History of Serous Ovarian Cancer: Defining the Target for Early Detection’. Edited by Paul D. P. Pharoah. *PLoS Medicine* 6 (7): e1000114. doi:10.1371/journal.pmed.1000114.
- 26 Kurman RJ and IM Shih, “The Origin and Pathogenesis of Epithelial Ovarian Cancer: A Proposed Unifying Theory,” *AMERICAN JOURNAL OF SURGICAL PATHOLOGY* 34, no. 3 (March 2010): 433-443.
- 27 Köbel, M., J. Madore, S. J. Ramus, B. A. Clarke, P. D. P. Pharoah, S. Deen, D. D. Bowtell, et al. 2014. ‘Evidence for a Time-Dependent Association between FOLR1 Expression and Survival from Ovarian Carcinoma: Implications for Clinical Testing. An Ovarian Tumour Tissue Analysis Consortium Study’. *British Journal of Cancer* 111 (12): 2297–2307. doi:10.1038/bjc.2014.567.

- 28 Buys SS, Partridge E, Black A, and et al. 2011. 'Effect of Screening on Ovarian Cancer Mortality: The Prostate, Lung, Colorectal and Ovarian (Plco) Cancer Screening Randomized Controlled Trial'. *JAMA* 305 (22): 2295–2303. doi:10.1001/jama.2011.766.
- 29 UK Collaborative Trial of Ovarian Cancer Screening Website: <http://www.ukctocs.org.uk/>
- 30 Sharma, A., M. Burnell, A. Gentry-Maharaj, S. Campbell, N. N. Amso, M. W. Seif, G. Fletcher, et al. 2013. 'Factors Affecting Visualization of Postmenopausal Ovaries: Descriptive Study from the Multicenter United Kingdom Collaborative Trial of Ovarian Cancer Screening (UKCTOCS)'. *Ultrasound in Obstetrics & Gynecology* 42 (4): 472–77. doi:10.1002/uog.12447.
- 31 Rayleigh, Lord. 'Investigations in Optics, with Special Reference to the Spectroscope'. *Monthly Notices of the Royal Astronomical Society* 40 (1 February 1880): 254. doi:10.1093/mnras/40.4.254.
- 32 Brunner, Eberhard. 'Ultrasound Front-End IC Considerations - UltrasoundFrontend.pdf'. *Ultrasound System Considerations and Their Impact on Front- End Components*, 2002. <http://www.analog.com/library/analogDialogue/archives/36-03/ultrasound/UltrasoundFrontend.pdf>.
- 33 Ultrasound Machine Reviews, News, and Advice: Ultrasound Reviews. 2016. Accessed January 5. <http://blog.gpsmed.com/ultrasound-reviews/>.
- 34 Wynd, Kimberly P., Hugh M. Smith, Adam K. Jacob, Laurence C. Torsher, Sandra L. Kopp, and James R. Hebl. 'Ultrasound Machine Comparison: An Evaluation of Ergonomic Design, Data Management, Ease of Use, and Image Quality'. *Regional Anesthesia and Pain Medicine* 34, no. 4 (July 2009): 349–56. doi:10.1097/AAP.0b013e3181ac9e5b.
- 35 Moran, Carmel M., Scott Inglis, and Stephen D. Pye. 2014. 'The Resolution Integral – a Tool for Characterising the Performance of Diagnostic Ultrasound Scanners'. *Ultrasound* 22 (1): 37–43. doi:10.1177/1742271X13518202.
- 36 Pye, S. D., W. Ellis, and T. MacGillivray. 2004. 'Medical Ultrasound: A New Metric of Performance for Grey-Scale Imaging'. *Journal of Physics: Conference Series* 1 (1): 187. doi:10.1088/1742-6596/1/1/041.
- 37 Haar T., G. *Ultrasound Bio-Effects and Safety Considerations*. Vol. 36. Karger Publishers, 2014. <http://www.karger.com/Article/Abstract/366233>.
- 38 Health, Center for Devices and Radiological. 'Medical Imaging - MRI (Magnetic Resonance Imaging)'. WebContent. Accessed 3 January 2016. <http://www.fda.gov/Radiation-EmittingProducts/RadiationEmittingProductsandProcedures/MedicalImaging/ucm200086.htm>.
- 39 Macdonald R., "Ultrasound and gynaecological cancer: current and potential future uses," *Ultrasound Journal of the British Medical Ultrasound Society* 18, no. 4 (November 2010): 18-175.
- 40 Govindan, Pramod, and Jafar Saniie. 'Processing Algorithms for Three-Dimensional Data Compression of Ultrasonic Radio Frequency Signals'. *IET Signal Processing* 9, no. 3 (16 April 2015): 267–76. doi:10.1049/iet-spr.2014.0186.

References

- 41 Ueng, Shyh-Kuang, Cho-Li Yen, Guan-Zhi Chen, Shyh-Kuang Ueng, Cho-Li Yen, and Guan-Zhi Chen. 'Ultrasound Image Enhancement Using Structure-Based Filtering, Ultrasound Image Enhancement Using Structure-Based Filtering'. *Computational and Mathematical Methods in Medicine, Computational and Mathematical Methods in Medicine* 2014, 2014 (19 June 2014): e758439. doi:10.1155/2014/758439, 10.1155/2014/758439.
- 42 Zhang Ju., Chen Wang, and Yun Cheng. 'Comparison of Despeckle Filters for Breast Ultrasound Images'. *Circuits, Systems, and Signal Processing* 34, no. 1 (18 June 2014): 185–208. doi:10.1007/s00034-014-9829-y.
- 43 Ploquin M., Adrian Basarab, and Denis Kouamé. 'Resolution Enhancement in Medical Ultrasound Imaging'. *Journal of Medical Imaging* 2, no. 1 (2015): 017001–017001. doi:10.1117/1.JMI.2.1.017001.
- 44 Yang S., Yaxin Sun, Yiguang Chen, and Licheng Jiao. 'Structural Similarity Regularized and Sparse Coding Based Super-Resolution for Medical Images'. *Biomedical Signal Processing and Control, Biomedical Image Restoration and Enhancement*, 7, no. 6 (November 2012): 579–90. doi:10.1016/j.bspc.2012.08.001.
- 45 HistoScanning™ by AMD » The Technology. 2016. Accessed April 20. <http://new.histoscanning.webfactional.com/prostate-histoscanningtm-tt/the-technology/>.
- 46 Lucidarme O., Jean-Paul Akakpo, Seth Granberg, Mario Sideri, Hanoch Levavi, Achim Schneider, Philippe Autier, Dror Nir, Harry Bleiberg, and for the Ovarian HistoScanning Clinical Study Group. 2010. 'A New Computer-Aided Diagnostic Tool for Non-Invasive Characterisation of Malignant Ovarian Masses: Results of a Multicentre Validation Study'. *European Radiology* 20 (8): 1822–30. doi:10.1007/s00330-010-1750-6.
- 47 Medison Inc. 'Accuvix XQ Technical Specification' v1.07 – www.medison.com
- 48 Grichnik, Katherine. 2012. 'Knobology'. February 6. <https://www.scahq.org/sca3/events/2012/echo/fundamentals/submissions/mon/Knobology%20-%20Grichnik.pdf>.
- 49 Ultrasound | Carestream. Accessed 10 January 2016. <http://www.carestream.com>.
- 50 Ultrasound Machine Comparison Table. Accessed 5 January 2016. <http://www.umiultrasound.com/compare-systems>.
- 51 Donald, I., J. Macvicar, and Tg Brown. 'Investigation of Abdominal Masses by Pulsed Ultrasound'. *Lancet* 1, no. JUN7 (1958): 1188–94.
- 52 Sanyal, Rupan, Benjamin Kraft, Lauren F. Alexander, Aimen Ismail, Mark E. Lockhart, and Michelle L. Robbin. 2016. 'Scanner-Based Protocol-Driven Ultrasound: An Effective Method to Improve Efficiency in an Ultrasound Department'. *American Journal of Roentgenology* 206 (4): 792–96.
- 53 Deming, W. Edwards, Joyce Orsini, and Diana Deming Cahill. 2012. *The Essential Deming: Leadership Principles from the Father of Quality*. 1 edition. McGraw-Hill Education.
- 54 NHS Institute for Innovation and Improvement Website – Quality and Value http://www.institute.nhs.uk/quality_and_value/lean_thinking/lean_six_sigma.html

- 55 Williams, Patricia M. 2001. 'Techniques for Root Cause Analysis'. Proceedings (Baylor University. Medical Center) 14 (2): 154–57.
- 56 The Economist. 2008. 'The Hawthorne Effect', November 3. <http://www.economist.com/node/12510632>.
- 57 Taguchi, Genichi, Subir Chowdhury, and Yuin Wu. 2004. Taguchi's Quality Engineering Handbook. 1 edition. Hoboken, N.J. : Livonia, Mich: Wiley-Interscience.
- 58 Bhote, Keki, and Adi Bhote. 2000. World Class Quality: Using Design of Experiments to Make It Happen. 2 Sub edition. New York: AMACOM.
- 59 Womack, James P., Daniel T. Jones, and Daniel Roos. 2007. The Machine That Changed the World: The Story of Lean Production-- Toyota's Secret Weapon in the Global Car Wars That Is Now Revolutionizing World Industry. Reprint edition. New York: Free Press.
- 60 University of Cambridge - Management Information Policy. 2016. 'Kanbans'. Accessed June 28. <http://www.ifm.eng.cam.ac.uk/research/dstools/kanbans/>.
- 61 da Silva, A. M., and M. C. C. Baranauskas. 2000. The ANDON System: Designing a CSCW Environment in a Lean Organization. Los Alamitos: Ieee Computer Soc.
- 62 Jones, Steven. 2016. Lean Six Sigma: The Ultimate Guide To Lean Six Sigma With Tools For Improving Quality And Speed! CreateSpace Independent Publishing Platform.
- 63 Smalley, Isao Kato Art. 2011. Toyota Kaizen Methods : Six Steps to Improvement. Boca Raton: Productivity Press.
- 64 Taleb, Nassim Nicholas. 2010. The Black Swan: Second Edition: The Impact of the Highly Improbable: With a New section: 'On Robustness and Fragility'. 2 edition. New York: Random House Trade Paperbacks.
- 65 Better Value in the NHS – Report Summary | The King's Fund. 2016. Accessed March 5. http://www.kingsfund.org.uk/publications/better-value-nhs/summary?gclid=CjwKEAiApOq2BRDoo8SVjZHV7TkSJABLe2iDoK8Wh-NHv8_CD7n5o82I71xAsgaOqJICLhnpL_oFzRoCVc7w_wcB.
- 66 Nicolay, C. R., S. Purkayastha, A. Greenhalgh, J. Benn, S. Chaturvedi, N. Phillips, and A. Darzi. 2012. 'Systematic Review of the Application of Quality Improvement Methodologies from the Manufacturing Industry to Surgical Healthcare'. British Journal of Surgery 99 (3): 324–35. doi:10.1002/bjs.7803.
- 67 Hughes, Ronda G. 2008. 'Tools and Strategies for Quality Improvement and Patient Safety'. In Patient Safety and Quality: An Evidence-Based Handbook for Nurses, edited by Ronda G. Hughes. Advances in Patient Safety. Rockville (MD): Agency for Healthcare Research and Quality (US). <http://www.ncbi.nlm.nih.gov/books/NBK2682/>.
- 68 Thor, Johan, Jonas Lundberg, Jakob Ask, Jesper Olsson, Cheryl Carli, Karin Pukk Härenstam, and Mats Brommels. 2007. 'Application of Statistical Process Control in Healthcare Improvement: Systematic Review'. Quality and Safety in Health Care 16 (5): 387–99. doi:10.1136/qshc.2006.022194.
- 69 Davidoff, Frank, paul Batalden, Greg ogrinc, and Susan Mooney. 2008. 'Publication Guidelines for Quality Improvement Studies in Healthcare: Evolution of the SQUIRE Project'. Canadian Journal of Diabetes, Canadian Journal of Diabetes, 32 (4): 281–89. doi:10.1016/S1499-2671(08)24008-6.

References

- 70 Kaplan, Heather C., Lloyd P. Provost, Craig M. Froehle, and Peter A. Margolis. 2012. 'The Model for Understanding Success in Quality (MUSIQ): Building a Theory of Context in Healthcare Quality Improvement'. *BMJ Quality & Safety* 21 (1): 13–20. doi:10.1136/bmjqs-2011-000010.
- 71 NHS Funding Shortfall Set to Widen While Social Care Financing “Unsustainable”, New Report Reveals | The Health Foundation. 2016. Accessed March 5. <http://www.health.org.uk/news/nhs-funding-shortfall-set-widen-while-social-care-financing-unsustainable-new-report-reveals>.
- 72 Mid Staffordshire NHS Foundation Trust Public Inquiry. 2016. Accessed March 5. <http://webarchive.nationalarchives.gov.uk/20150407084003/http://www.midstaffspublicinquiry.com/report>.
- 73 Morecambe Bay Investigation - GOV.UK. 2016. Accessed March 5. <https://www.gov.uk/government/organisations/morecambe-bay-investigation>.
- 74 Toyota Production System. 2016. Toyota Australia. Accessed March 5. <http://www.toyota.com.au/toyota/company/operations/toyota-production-system>.
- 75 Boeing Frontiers Online. 2016. Accessed March 5. http://www.boeing.com/news/frontiers/archive/2003/july/i_ca1.html.
- 76 NHS England » The NHS Belongs to the People: A Call to Action. 2016. Accessed March 5. <https://www.england.nhs.uk/2013/07/call-to-action/>.
- 77 Quality Improvement Hub. 2015. December 15. <http://www.qihub.scot.nhs.uk/home.aspx>.
- 78 Improvement, 1000 Lives. 2016. '1000 Lives Improvement - Improving Quality Together'. Accessed March 5. <http://www.1000livesplus.wales.nhs.uk/iqt>.
- 79 Northern Ireland - Regulation and Quality Improvement Authority (RQIA). Text. May 6. <http://www.rqia.org.uk/home/index.cfm>.
- 80 Quality 2020 - a Ten Year Strategy to Protect and Improve Quality in Health and Social Care in Northern Ireland. 2015. DHSSPS. August 24. <https://www.dhsspsni.gov.uk/publications/quality-2020-ten-year-strategy-protect-and-improve-quality-health-and-social-care>.
- 81 NHS NI - Safety and Quality Standards. 2015. DHSSPS. August 12. <https://www.dhsspsni.gov.uk/topics/safety-and-quality-standards>.
- 82 NHS England » Sustainable Improvement Team. 2016. Accessed March 5. <https://www.england.nhs.uk/ourwork/qual-clin-lead/nhsiq/>.
- 83 NHS IQ. 2016. Accessed March 6. <http://www.nhsiq.nhs.uk/>.
- 84 Quality Improvement: Theory and Practice in Healthcare - NHS Institute for Innovation and Improvement. 2016. Accessed March 4. http://www.institute.nhs.uk/service_transformation/quality_improvement/quality_improvement%3A_theory_and_practice_in_healthcare.html.
- 85 Health and Social Care Act 2012. 2016. Accessed March 5. <http://www.legislation.gov.uk/ukpga/2012/7/section/23/enacted>.
- 86 NICE -Ovarian Cancer | Guidance and Guidelines | NICE. 2016. Accessed March 14. <https://www.nice.org.uk/guidance/qs18>.

- 87 NICE - Audit and Service Improvement | Into Practice | What We Do | About | NICE. 2016. CorporatePage. Accessed March 14. <https://www.nice.org.uk/about/what-we-do/into-practice/audit-and-service-improvement>.
- 88 NICE - Into Practice Guide | 1 Introduction and Background | Guidance and Guidelines | NICE. 2016. Accessed March 14. <https://www.nice.org.uk/article/pg1/chapter/1%20Introduction%20and%20background>.
- 89 C4EO - Change Models - Introduction. 2016. Accessed March 14. <http://archive.c4eo.org.uk/changemodels/>.
- 90 High Quality Care for All: The NHS Next Stage Review Final Report. 2008. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/228836/7432
- 91 Quality in the New Health System – Maintaining and Improving Quality from April 2013 - Publications - GOV.UK. 2016. Accessed March 5. <https://www.gov.uk/government/publications/quality-in-the-new-health-system-maintaining-and-improving-quality-from-april-2013>.
- 92 NHS England » Five Year Forward View. 2016. Accessed March 5. <https://www.england.nhs.uk/ourwork/futurenhs/>.
- 93 NHS England » Agendas, Minutes and Papers. 2016. Accessed March 6. <https://www.england.nhs.uk/ourwork/part-rel/nqb/agen-mins-meet/>.
- 94 Quality Improvement Programme | The Royal College Of Radiologists. 2016. Accessed March 5. <https://www.rcr.ac.uk/clinical-radiology/audit-and-qi/quality-improvement-programme>.
- 95 Healthcare Quality Improvement Partnership - Data.gov.uk. 2016. Accessed March 5. <https://data.gov.uk/publisher/healthcare-quality-improvement-partnership>.
- 96 HQIP-Guide-to-Using-Quality-Improvement-Tools-to-Drive-CA-2011.pdf. 2016. Accessed March 14. <http://www.hqip.org.uk/public/cms/253/625/19/193/HQIP-Guide-to-Using-Quality-Improvement-Tools-to-Drive-CA-2011.pdf?realName=NIgOW4.pdf>.
- 97 RCOG - Delivery of Ultrasound Training. 2016. Royal College of Obstetricians & Gynaecologists. Accessed March 16. <https://www.rcog.org.uk/en/careers-training/specialty-training-curriculum/ultrasound-training/introduction/delivery-of-ultrasound-training/>.
- 98 Thomson, Nigel, and Audrey Paterson. 2014. 'Sonographer Registration in the United Kingdom – a Review of the Current Situation'. *Ultrasound* 22 (1): 52–56. doi:10.1177/1742271X13517381.
- 99 RCR Standards for the Provision of an Ultrasound Service. 2016. Accessed March 17. [https://www.rcr.ac.uk/sites/default/files/documents/BFCR\(14\)17_Standards_ultrasound.pdf](https://www.rcr.ac.uk/sites/default/files/documents/BFCR(14)17_Standards_ultrasound.pdf).
- 100 Regulatory Framework - European Commission. 2016. Accessed March 18. http://ec.europa.eu/growth/sectors/medical-devices/regulatory-framework/index_en.htm.
- 101 NHSLA Risk Management Standards 2012-13 - NHS LA Risk Management Standards 2013-14.pdf. 2016. Accessed March 18. <http://www.nhsla.com/Safety/Documents/NHS%20LA%20Risk%20Management%20Standards%202013-14.pdf>.

References

- 102 Dudley, Nick, Stephen Russell, Barry Ward, Peter Hoskins, and BMUS QA working Party. 2014. 'The BMUS Guidelines for Regular Quality Assurance Testing of Ultrasound Scanners'. *Ultrasound* 22 (1): 6–7. doi:10.1177/1742271X13515557.
- 103 Russell, Stephen. 2014. 'Ultrasound Quality Assurance and Equipment Governance'. *Ultrasound* 22 (1): 66–69. doi:10.1177/1742271X13517694.
- 104 Dudley, N. J., and N. M. Gibson. 2014. 'Early Experience with Automated B-Mode Quality Assurance Tests'. *Ultrasound* 22 (1): 15–20. doi:10.1177/1742271X13516896.
- 105 RCR - Quality Improvement Programme | The Royal College Of Radiologists. 2016. Accessed March 5. <https://www.rcr.ac.uk/clinical-radiology/audit-and-qi/quality-improvement-programme>.
- 106 Dudley, N. J., K. Griffith, G. Houldsworth, M. Holloway, and M. A. Dunn. 2001. 'A Review of Two Alternative Ultrasound Quality Assurance Programmes'. *European Journal of Ultrasound* 12 (3): 233–45. doi:10.1016/S0929-8266(00)00119-1.
- 107 Miyague, André Hadyme, Fernando Marum Mauad, Wellington de Paula Martins, Augusto César Garcia Benedetti, Ana Elizabeth Gomes de Melo Tavares Ferreira, Francisco Mauad-Filho, André Hadyme Miyague, et al. 2015. 'Ultrasound Scan as a Potential Source of Nosocomial and Cross-Infection: A Literature Review'. *Radiologia Brasileira* 48 (5): 319–23. doi:10.1590/0100-3984.2014.0002.
- 108 Cantin, Peter, and Karen Knapp. 2014. 'Diagnostic Image Quality in Gynaecological Ultrasound: Who Should Measure It, What Should We Measure and How?' *Ultrasound* 22 (1): 44–51. doi:10.1177/1742271X13511242.
- 109 BMUS - Professional Guidance. 2016. Accessed March 16. <https://www.bmus.org/policies-statements-guidelines/professional-guidance/>.
- 110 Ovarian Cancer: Recognition and Initial Management | Guidance and Guidelines | NICE. 2016. Accessed March 21. <https://www.nice.org.uk/guidance/cg122>.
- 111 Sayasneh, Ahmad, Jeroen Kaijser, Jessica Preisler, Susanne Johnson, Catriona Stalder, Richard Husicka, Sharmistha Guha, et al. 2013. 'A Multicenter Prospective External Validation of the Diagnostic Performance of IOTA Simple Descriptors and Rules to Characterize Ovarian Masses'. *Gynecologic Oncology* 130 (1): 140–46. doi:10.1016/j.ygyno.2013.04.003.
- 112 RCOG Green-Top Guidelines. 2016. Royal College of Obstetricians & Gynaecologists. Accessed March 20. <https://www.rcog.org.uk/en/guidelines-research-services/guidelines/>.
- 113 Fetal Anomaly Screening: Programme Overview - Detailed Guidance - GOV.UK. Accessed March 24. <https://www.gov.uk/guidance/fetal-anomaly-screening-programme-overview>.
- 114 Down's Syndrome Screening: Risk Calculation Software Requirements - Publications - GOV.UK. 2016. Accessed March 21. <https://www.gov.uk/government/publications/downs-syndrome-screening-risk-calculation-software-requirements>.
- 115 Twin pregnancies: second trimester screening for Down's syndrome (T2) . 2016. Accessed March 25. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/461527/FA_SP_Quad_screening_twins_final_Sept_2015_print_ready.pdf.

- 116 Nicolaides, K. H., G. Azar, D. Byrne, C. Mansur, and K. Marks. 1992. 'Fetal Nuchal Translucency: Ultrasound Screening for Chromosomal Defects in First Trimester of Pregnancy.' *BMJ* 304 (6831): 867–69. doi:10.1136/bmj.304.6831.867.
- 117 Fetal Anomaly Screening: Ultrasound Practitioners Handbook - Publications - GOV.UK. 2016. Accessed March 21. <https://www.gov.uk/government/publications/fetal-anomaly-screening-ultrasound-practitioners-handbook>.
- 118 Antenatal and Newborn Screening: Quality Assurance Procedures - Publications - GOV.UK. 2016. Accessed March 24. <https://www.gov.uk/government/publications/antenatal-and-newborn-screening-quality-assurance-procedures>.
- 119 Fetal Anomaly Screening Programme: Standards - Publications - GOV.UK. 2016. Accessed March 21. <https://www.gov.uk/government/publications/fetal-anomaly-screening-programme-standards>.
- 120 Fetal Medicine Foundation. 2016. Accessed March 25. <https://fetalmedicine.org/>.
- 121 Palomaki, Glenn E., Jo Ellen S. Lee, Jacob A. Canick, Geraldine A. McDowell, and Alan E. Donnemfeld. 2009. 'Technical Standards and Guidelines: Prenatal Screening for Down Syndrome That Includes First-Trimester Biochemistry And/or Ultrasound Measurements'. *Genetics in Medicine* 11 (9): 669–81. doi:10.1097/GIM.0b013e3181ad5246.
- 122 Hynek, Martin, Dagmar Smetanová, David Stejskal, and Jana Zvárová. 2014. 'Exponentially Weighted Moving Average Chart as a Suitable Tool for Nuchal Translucency Quality Review'. *Prenatal Diagnosis* 34 (4): 367–76. doi:10.1002/pd.4314.
- 123 NSC Recommendations on Ovarian Cancer Screening. 2016. Accessed March 26. <http://legacy.screening.nhs.uk/ovariancancer>.
- 124 Weissfeld, J. L., R. M. Fagerstrom, and B. O'Brien. 2000. 'Quality Control of Cancer Screening Examination Procedures in the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial'. *Controlled Clinical Trials* 21 (6): 390S – 399S. doi:10.1016/S0197-2456(00)00094-5.
- 125 Salomon, L. J., M. Nassar, J. P. Bernard, Y. Ville, and A. Fauconnier. 2009. 'A Score-Based Method to Improve the Quality of Emergency Gynaecological Ultrasound Examination'. *European Journal of Obstetrics & Gynecology and Reproductive Biology* 143 (2): 116–20. doi:10.1016/j.ejogrb.2008.12.003.
- 126 Bredella, Miriam A., Vickie A. Feldstein, Roy A. Filly, Ruth B. Goldstein, Peter W. Callen, and Harry K. Genant. 2000. 'Measurement of Endometrial Thickness at US in Multicenter Drug Trials: Value of Central Quality Assurance Reading'. *Radiology* 217 (2): 516–20. doi:10.1148/radiology.217.2.r00nv34516.
- 127 Burger, M., T. Weber-Rössler, and M. Willmann. 1997. 'Measurement of the Pregnant Cervix by Transvaginal Sonography: An Interobserver Study and New Standards to Improve the Interobserver Variability'. *Ultrasound in Obstetrics and Gynecology* 9 (3): 188–93. doi:10.1046/j.1469-0705.1997.09030188.x.
- 128 Gohagan, John K., Barbara O'Brien, Marsha A. Hasson, Keith D. Umbel, Beth Bridgeman, Barnett S. Kramer, Douglas Reding, et al. 2015. 'Comprehensive Quality Management (CQM) in the PLCO Trial'. *Reviews on Recent Clinical Trials* 10 (3): 223–32.

References

- 129 Hasson, Marsha A., Richard M. Fagerstrom, Dalia C. Kahane, Judith H. Walsh, Max H. Myers, Clifford Caughman, Blaine Wenzel, Juline C. Haralson, Lynn M. Flickinger, and Louisa M. Turner. 2000. 'Design and Evolution of the Data Management Systems in the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial'. *Controlled Clinical Trials* 21 (6, Supplement 1): 329S – 348S. doi:10.1016/S0197-2456(00)00100-8.
- 130 Moore, Stephen M., David S. Gierada, Kenneth W. Clark, and G. James Blaine. 2005. 'Image Quality Assurance in the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial Network of the National Lung Screening Trial'. *Journal of Digital Imaging* 18 (3): 242–50. doi:10.1007/s10278-005-5153-1.
- 131 Gohagan, J., Philip C. Prorok. 2015. 'The PLCO Cancer Screening Trial: Background, Goals, Organization, Operations, Results'. *Reviews on Recent Clinical Trials* 10 (3). doi:10.2174/1574887110666150730123004.
- 132 Mabie, Jerome, Tom Riley, Pamela M. Marcus, Amanda Black, Heather Rozjabek, Kelly Yu, Michael Young, et al. 2015. 'Data Processing and Analytic Support in the PLCO Cancer Screening Trial'. *Reviews on Recent Clinical Trials* 10 (3): 233–37.
- 133 Yourdon, Edward, and Larry L. Constantine. 1991. *Structured Design: Fundamentals of a Discipline of Computer Programme and Systems Design*. 2nd Revised edition edition. Ashgate Publishing Limited.
- 134 Philips Healthcare | DICOM - WebView Systems. 2016. Philips. Accessed April 12. <http://www.philips.co.uk/healthcare/resources/support-documentation/dicom-web-viewing>.
- 135 Gentry-Maharaj, A., A. Sharma, M. Burnell, A. Ryan, N. N. Amso, M. W. Seif, G. Turner, et al. 2013. 'Acceptance of Transvaginal Sonography by Postmenopausal Women Participating in the United Kingdom Collaborative Trial of Ovarian Cancer Screening'. *Ultrasound in Obstetrics & Gynecology* 41 (1): 73–79. doi:10.1002/uog.12262.
- 136 Stott, Will, and James Newkirk. 2007. *Visual Studio Team System: Better Software Development for Agile Teams*. 1 edition. Addison-Wesley Professional.
- 137 Watanabe, Katsuaki. 2007. 'Lessons from Toyota's Long Drive'. *Harvard Business Review* 85 (7-8): 74 – +.
- 138 OSPACS. 2016. CodePlex. Accessed May 30. <http://ospacs.codeplex.com/Wikipage?ProjectName=ospacs>.
- 139 Standard Floppy Disk Formats Supported by MS-DOS. 2016. Accessed May 30. <https://support.microsoft.com/en-us/kb/75131>.
- 140 ezDICOM Download | SourceForge.net. 2016. Accessed May 30. <https://sourceforge.net/projects/ezdicom/>.
- 141 Stott, Will. 2013. 'IfWH Biobank – a Feasibility Study. May 2013.'
- 142 Red Gate Software. 2016. 'Redgate Software - Tools For SQL Server, .NET, & Oracle'. Accessed May 30. <http://www.red-gate.com/>.
- 143 Ambler, Scott J., and Pramod J. Sadalage. 2006. *Refactoring Databases: Evolutionary Database Design*. 1 edition. Upper Saddle River, NJ: Addison-Wesley Professional.

- 144 Labvantage | LIMS, Biobanking and Quality ELN | LabVantage | Increase Lab Efficiency. 2016. Accessed May 30. <http://www.labvantage.com/>.
- 145 Bika Lab Systems. 2016. Accessed May 30. <http://www.bikalabs.com/article/newsevents/bikabiobank>.
- 146 OpenSpecimen Streamlines Management of Biospecimen Data. 2016. Opensource.com. Accessed May 30. <https://opensource.com/health/15/3/OpenSpecimen-biobank-informatics>.
- 147 OpenELIS. 2016. Accessed May 30. <http://www.openelis-foundation.org/>.
- 148 LabKey – Open-Source Freedom, Enterprise Quality. 2016. Accessed May 30. <https://www.labkey.com/>.
- 149 Concept - Open-LIMS - The Open-Source Laboratory Information Management System. 2016. Accessed May 30. <http://open-lims.org/concept.html>.
- 150 OpenClinica - Electronic Data Capture for Clinical Research. 2016. OpenClinica. Accessed May 30. <https://www.openclinica.com/>.
- 151 EDC - ClinCapture®. 2016. ClinCapture - Clinical Data Management Software. Accessed May 30. <http://www.clincapture.com/products/clincapture>.
- 152 Microsoft Excel | Spreadsheet Software. 2016. Accessed May 27. <https://products.office.com/en-gb/excel>.
- 153 STDU Viewer. 2016. Accessed May 27. <http://www.stdutility.com/stdviewer.html>.
- 154 Visual Studio - Microsoft Developer Tools. 2016. Accessed May 9. <https://www.visualstudio.com/>.
- 155 Microsoft Access Database Engine 2010 Redistributable. 2016. Microsoft Download Center. Accessed May 28. <https://www.microsoft.com/en-gb/download/details.aspx?id=13255>.
- 156 R: The R Project for Statistical Computing. 2016. Accessed May 29. <https://www.r-project.org/>.
- 157 RUnit - A Unit Test Framework for R. 2016. Accessed May 29. <https://cran.r-project.org/web/packages/RUnit/vignettes/RUnit.pdf>.
- 158 MATLAB - MathWorks - MathWorks United Kingdom. 2016. Accessed May 29. <http://uk.mathworks.com/products/matlab/>.
- 159 TinyXml Main Page. 2016. Accessed May 23. <http://www.grinninglizard.com/tinyxml/>.
- 160 Doubilet, P., C. Benson. 2011. Atlas of Ultrasound in Obstetrics and Gynecology: A Multimedia Reference. Second edition. Philadelphia: LWW.
- 161 IBM Archives: System/360 Announcement. 2003. TS200. January 23. http://www-03.ibm.com/ibm/history/exhibits/mainframe/mainframe_PR360.html.
- 162 Sessions, Roger. 2000. COM+ and the Battle for the Middle Tier. 1 edition. New York: Wiley.
- 163 IBM MQ. 2016. Accessed May 3. <http://www-03.ibm.com/software/products/en/ibm-mq>.

References

- 164 Message Queuing (MSMQ). 2016. Accessed May 3. [https://msdn.microsoft.com/en-us/library/ms711472\(v=vs.85\).aspx](https://msdn.microsoft.com/en-us/library/ms711472(v=vs.85).aspx).
- 165 Google Research Publication: MapReduce. 2016. Accessed May 5. <http://research.google.com/archive/mapreduce.html>.
- 166 Welcome to Apache™ Hadoop®! 2016. Accessed May 5. <https://hadoop.apache.org/>.
- 167 Amazon Web Services (AWS) - Cloud Computing Services. 2016. Amazon Web Services, Inc. Accessed May 5. <http://aws.amazon.com/>.
- 168 Microsoft Azure: Cloud Computing Platform and Services. 2016. Accessed May 5. <https://azure.microsoft.com/en-gb/>.
- 169 Challenges in Irreproducible Research : Nature News & Comment. 2016. Accessed May 6. <http://www.nature.com.libproxy.ucl.ac.uk/news/reproducibility-1.17552>.
- 170 Windows Workflow Foundation Overview. 2016. Accessed April 21. [https://msdn.microsoft.com/en-us/library/ms734631\(v=vs.90\).aspx](https://msdn.microsoft.com/en-us/library/ms734631(v=vs.90).aspx).
- 171 Taverna - Apache Incubator. 2016. Accessed April 21. <http://incubator.apache.org/projects/taverna.html>.
- 172 VisTrailsWiki. 2016. Accessed April 21. http://www.vistrails.org/index.php/Main_Page.
- 173 Cui, Zaixu, Chenxi Zhao, and Gaolang Gong. 2015. 'Parallel Workflow Tools to Facilitate Human Brain MRI Post-Processing'. *Frontiers in Neuroscience* 9: 171. doi:10.3389/fnins.2015.00171.
- 174 Gorgolewski, Krzysztof, Christopher D. Burns, Cindee Madison, Dav Clark, Yaroslav O. Halchenko, Michael L. Waskom, and Satrajit S. Ghosh. 2011. 'Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python'. *Frontiers in Neuroinformatics* 5: 13. doi:10.3389/fninf.2011.00013.
- 175 Rex, David E, Jeffrey Q Ma, and Arthur W Toga. 2003. 'The LONI Pipeline Processing Environment'. *NeuroImage* 19 (3): 1033–48. doi:10.1016/S1053-8119(03)00185-X.
- 176 Li, Bo, Frederick Bryan, and Bennett A. Landman. 2012. 'Next Generation of the Java Image Science Toolkit (JIST): Visualization and Validation'. *The Insight Journal* 2012 (August): 1–16.
- 177 What Are The Current Challenges That Users Face When Using Nipype? 2016. Accessed April 21. <https://neurostars.org/p/73/>.
- 178 CLI Wrapper Automatization. 2016. Accessed April 21. <https://neurostars.org/p/159/#160>.
- 179 Open Grid Scheduler: The Official Open Source Grid Engine. 2016. Accessed April 21. <http://gridscheduler.sourceforge.net/>.
- 180 Parallel Computing Toolbox - MATLAB - MathWorks United Kingdom. 2016. Accessed April 21. http://uk.mathworks.com/products/parallel-computing/index.html?s_tid=gn_loc_drop.
- 181 Stroustrup: The C++ Programming Language (4th Edition). 2016. Accessed May 9. <http://www.stroustrup.com/4th.html>.
- 182 Arnold, Ken, James Gosling, and David Holmes. 2005. *The Java Programming Language, 4th Edition*. 4 edition. Upper Saddle River, NJ: Addison-Wesley Professional.

- 183 Hejlsberg, Anders, Mads Torgersen, Scott Wiltamuth, and Peter Golde. 2010. *The C# Programming Language*. 4 edition. Upper Saddle River, NJ: Addison-Wesley Professional.
- 184 Dynamic-Link Libraries (Windows). 2016. Accessed May 9. [https://msdn.microsoft.com/en-us/library/windows/desktop/ms682589\(v=vs.85\).aspx](https://msdn.microsoft.com/en-us/library/windows/desktop/ms682589(v=vs.85).aspx).
- 185 Box, Don. 1998. *Essential COM*. 1 edition. Reading, Mass: Addison-Wesley Professional.
- 186 Using the MIDL Compiler (Windows). 2016. Accessed May 10. [https://msdn.microsoft.com/en-us/library/windows/desktop/aa367300\(v=vs.85\).aspx](https://msdn.microsoft.com/en-us/library/windows/desktop/aa367300(v=vs.85).aspx).
- 187 Martin, Robert C. 2002. *Agile Software Development, Principles, Patterns, and Practices*. 1st edition. Upper Saddle River, N.J: Pearson.
- 188 GetLastError Function (Windows). 2016. Accessed May 18. [https://msdn.microsoft.com/en-us/library/windows/desktop/ms679360\(v=vs.85\).aspx](https://msdn.microsoft.com/en-us/library/windows/desktop/ms679360(v=vs.85).aspx).
- 189 OpenCV | OpenCV. 2016. Accessed May 18. <http://opencv.org/>.
- 190 Intel® Integrated Performance Primitives (Intel® IPP) | Intel® Software. 2016. Accessed May 18. <https://software.intel.com/en-us/intel-ipp>.
- 191 CUDA | OpenCV. 2016. Accessed May 18. <http://opencv.org/platforms/cuda.html>.
- 192 OpenCL - The Open Standard for Parallel Programming of Heterogeneous Systems. 2016. The Khronos Group. Accessed May 18. <https://www.khronos.org/api/ocl/>.
- 193 MxPlugin. 2016. CodePlex. Accessed May 21. <http://mxplugin.codeplex.com/Wikipage?ProjectName=mxplugin>.
- 194 Smart Pointers (Modern C++). 2016. Accessed May 21. <https://msdn.microsoft.com/en-us/library/hh279674.aspx>.
- 195 WiX Toolset. 2016. WiX Toolset. Accessed May 22. <http://wixtoolset.org//index.html>.
- 196 Gamma, Erich, Richard Helm, Ralph Johnson, and John Vlissides. 1994. *Design Patterns: Elements of Reusable Object-Oriented Software*. 1 edition. Addison-Wesley Professional.
- 197 Fowler, Martin. 2010. *Domain-Specific Languages*. 1 edition. Upper Saddle River, NJ: Addison-Wesley Professional. Page 27
- 198 Bliki: LanguageWorkbench. 2016. Martinfowler.com. Accessed May 26. <http://martinfowler.com/bliki/LanguageWorkbench.html>.
- 199 BCGSoft | BCGSuite for MFC - Product Overview. 2016. Accessed May 26. <https://www.bcgsoft.com/bcgsuite.htm>.
- 200 Stott, W., A. Franchini, C. Jones, A. Ryan, S. Campbell, A. Gentry-Maharaj, G. Fletcher, J. Kalsi, I. Jacobs, U. Menon. 2016. 'Software Tool for quality control of Transvaginal Sonography (TVS) of postmenopausal women'. *Ultrasound in Obstetrics & Gynecology*. Not accepted or published.
- 201 Santosh, K. C., Naved Alam, Partha Pratim Roy, Laurent Wendling, Sameer Antani, and George R. Thoma. 2016. 'A Simple and Efficient Arrowhead Detection Technique in Biomedical Images'. *International Journal of Pattern Recognition and Artificial Intelligence* 30 (5): 1657002. doi:10.1142/S0218001416570020.

References

- 202 Dori, D., and Wenyin Liu. 1999. 'Sparse Pixel Vectorization: An Algorithm and Its Performance Evaluation'. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21 (3): 202–15. doi:10.1109/34.754586.
- 203 Dori, D., and Liu Wenyin. 1999. 'Automated CAD Conversion with the Machine Drawing Understanding System: Concepts, Algorithms, and Performance'. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 29 (4): 411–16. doi:10.1109/3468.769761.
- 204 Cheng, Beibei, R. Joe Stanley, Soumya De, Sameer Antani, and George R. Thoma. 2011. 'Automatic Detection of Arrow Annotation Overlays in Biomedical Images'. *International Journal of Healthcare Information Systems and Informatics* 6 (4): 23–41. doi:10.4018/jhisi.2011100102.
- 205 Daekeun, Y., Matthew Simpson, Sameer Antani, Dina Demner-Fushman, and George R. Thoma. 2013. 'A Robust Pointer Segmentation in Biomedical Images toward Building a Visual Ontology for Biomedical Article Retrieval'. In , 8658:86580Q–86580Q–9. doi:10.1117/12.2005934.
- 206 Sharma, A., S. Apostolidou, M. Burnell, S. Campbell, M. Habib, A. Gentry-Maharaj, N. Amso, et al. 2012. 'Risk of Epithelial Ovarian Cancer in Asymptomatic Women with Ultrasound-Detected Ovarian Masses: A Prospective Cohort Study within the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS)'. *Ultrasound in Obstetrics & Gynecology* 40 (3): 338–44. doi:10.1002/uog.12270.
- 207 Partridge, Edward E., Robert T. Greenlee, Thomas L. Riley, John Commins, Lawrence Ragard, Jian-Lun Xu, Sandra S. Buys, Philip C. Prorok, and Mona N. Fouad. 2013. 'Assessing the Risk of Ovarian Malignancy in Asymptomatic Women With Abnormal CA 125 and Transvaginal Ultrasound Scans in the Prostate, Lung, Colorectal, and Ovarian Screening Trial'. *Obstetrics & Gynecology* 121 (1): 25–31. doi:10.1097/AOG.0b013e3182755e14.
- 208 Partridge, Edward, Aimee R. Kreimer, Robert T. Greenlee, Craig Williams, Jian-Lun Xu, Timothy R. Church, Bruce Kessel, et al. 2009. 'Results From Four Rounds of Ovarian Cancer Screening in a Randomized Trial'. *Obstetrics & Gynecology* 113 (4): 775–82. doi:10.1097/AOG.0b013e31819cda77.
- 209 Stott, W., A. Ryan, A. Gentry-Maharaj, G. Fletcher, M. Burnell, N. Asmo, M. Seif, A. Ferguson, G. Turner, C. Brunell, K. Ford, R. Ranger, S. Campbell, C. Jones, I. Jacobs, U. Menon. 'Audit of transvaginal sonography (TVS) by sonographers from the United Kingdom Collaborative Trial of Ovarian Cancer Screening (UKCTOCS)' (not submitted, accepted or published).
- 210 Landis, J. Richard, and Gary G. Koch. 1977. 'The Measurement of Observer Agreement for Categorical Data'. *Biometrics* 33 (1): 159. doi:10.2307/2529310.
- 211 Viera, A. J., and J. M. Garrett. 2005. 'Understanding Interobserver Agreement: The Kappa Statistic'. *Family Medicine* 37 (5): 360–63.
- 212 Data Analysis and Statistical Software | Stata. 2016. Accessed June 15. <http://www.stata.com/>.
- 213 Zwick, R. 1988. 'Another Look at Interrater Agreement'. *Psychological Bulletin* 103 (3): 374–78. doi:10.1037//0033-2909.103.3.374.
- 214 Parpia, Sameer, John J. Koval, and Allan Donner. 2013. 'Evaluation of Confidence Intervals for the Kappa Statistic When the Assumption of Marginal Homogeneity Is Violated'. *Computational Statistics* 28 (6): 2709–18. doi:10.1007/s00180-013-0424-7.

- 215 Bodelon, Clara, Ruth M. Pfeiffer, Sandra S. Buys, Amanda Black, and Mark E. Sherman. 'Analysis of Serial Ovarian Volume Measurements and Incidence of Ovarian Cancer: Implications for Pathogenesis'. *Journal of the National Cancer Institute* 106, no. 10 (10 January 2014): dju262. doi:10.1093/jnci/dju262.
- 216 Van Nagell, John Rensselaer, Rachel Ware Miller, Christopher P. DeSimone, Frederick R. Ueland, Iwona Podzielinski, Scott T. Goodrich, Jeff W. Elder, Bin Huang, Richard J. Kryscio, and Edward John Pavlik. 'Long-Term Survival of Women With Epithelial Ovarian Cancer Detected by Ultrasonographic Screening'. *Obstetrics & Gynecology* 118, no. 6 (December 2011): 1212–21. doi:10.1097/AOG.0b013e318238d030.
- 217 Prorok, P. C., G. L. Andriole, R. S. Bresalier, S. S. Buys, D. Chia, E. D. Crawford, R. Fogel, et al. 'Design of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial'. *Controlled Clinical Trials* 21, no. 6 Suppl (December 2000): 273S – 309S.
- 218 Coleman, B G, P H Arger, K Grumbach, M K Menard, M C Mintz, K S Allen, R L Arenson, and K A Lamon. "Transvaginal and Transabdominal Sonography: Prospective Comparison." *Radiology* 168, no. 3 (September 1988): 639–43.
- 219 Fishman, D, L Cohen, S Blank, L Shulman, D Singh, K Bozorgi, R Tamura, I Timortritsch, and P Schwartz. "The Role of Ultrasound Evaluation in the Detection of Early-Stage Epithelial Ovarian Cancer." *American Journal of Obstetrics and Gynecology* 192, no. 4 (April 2005): 1214–21. doi:10.1016/j.ajog.2005.01.041.
- 220 Sherman, Mark E., James V. Lacey, Sandra S. Buys, Douglas J. Reding, Christine D. Berg, Craig Williams, and Patricia Hartge. "Ovarian Volume: Determinants and Associations with Cancer among Postmenopausal Women." *Cancer Epidemiology Biomarkers & Prevention* 15, no. 8 (2006): 1550–54. doi:10.1158/1055-9965.EPI-05-0847.
- 221 Healy, David L., Robin Bell, David M. Robertson, Tom Jobling, Martin K. Oehler, Andrew Edwards, Paul Shekleton, et al. "Ovarian Status in Healthy Postmenopausal Women." *Menopause* 15, no. 6 (December 2008): 1109–14. doi:10.1097/gme.0b013e31816ddb6f.
- 222 Stott, W ., S. Campbell, A. Franchini, O. Blyuss, A. Ryan, C. Jones, A. Gentry-Maharaj, G. Fletcher, J. Kalsi, S. Skates, M. Palmer, I. Jacobs, U. Menon. 2016. 'Quality Control of transvaginal sonography (TVS) using images from the United Kingdom Collaborative Trial of Ovarian Cancer Screening (UKCTOCS)'. *Ultrasound in Obstetrics and Gynecology*. (Submitted)
- 223 Khazendar, S., A. Sayasneh, H. Al-Assam, H. Du, J. Kaijser, L. Ferrara, D. Timmerman, S. Jassim, and T. Bourne. 2015. 'Automated Characterisation of Ultrasound Images of Ovarian Tumours: The Diagnostic Accuracy of a Support Vector Machine and Image Processing with a Local Binary Pattern Operator'. *Facts, Views & Vision in ObGyn* 7 (1): 7–15.
- 224 Khazendar, S., H. Al-Assam, H. Du, S. Jassim, A. Sayasneh, T. Bourne, J. Kaijser, and D. Timmerman. 2014. 'Automated Classification of Static Ultrasound Images of Ovarian Tumours Based on Decision Level Fusion'. In *Computer Science and Electronic Engineering Conference (CEEC), 2014 6th*, 148–53. doi:10.1109/CEEC.2014.6958571.
- 225 Kumar, H. Prasanna, and S. Srinivasan. 2015. 'Fast Automatic Segmentation of Polycystic Ovary in Ultrasound Images Using Improved Chan-Vase with Split-Bregman Optimization'. *Journal of Medical Imaging and Health Informatics* 5 (1): 57–62. doi:10.1166/jmihi.2015.1355.

References

- 226 Classification of Ovary Abnormality Using the Probabilistic Neural Network (PNN). 2016. Accessed June 13. <http://content.iospress.com.libproxy.ucl.ac.uk/download/technology-and-health-care/thc00863?id=technology-and-health-care%2Fthc00863>.
- 227 Chan, T. F., and L. A. Vese. 2001. 'Active Contours without Edges'. *IEEE Transactions on Image Processing* 10 (2): 266–77. doi:10.1109/83.902291.
- 228 Kass, Michael, Andrew Witkin, and Demetri Terzopoulos. 1988. 'Snakes: Active Contour Models'. *International Journal of Computer Vision* 1 (4): 321–31. doi:10.1007/BF00133570.
- 229 Reddy, G. Raghobham, K. Ramudu, P. Yugander, and R. Rameshwar Rao. 2011. 'Fast Global Region Based Minimization of Satellite and Medical Imagery with Geometric Active Contour and Level Set Evolution on Noisy Images'. In *2011 IEEE Recent Advances in Intelligent Computational Systems (RAICS)*, 696–700. doi:10.1109/RAICS.2011.6069400.
- 230 Bodelon, Clara, Ruth M. Pfeiffer, Sandra S. Buys, Amanda Black, and Mark E. Sherman. 'Analysis of Serial Ovarian Volume Measurements and Incidence of Ovarian Cancer: Implications for Pathogenesis'. *Journal of the National Cancer Institute* 106, no. 10 (10 January 2014): dju262. doi:10.1093/jnci/dju262.
- 231 Van Nagell Jr., J. R., P. D. DePriest, M. B. Reedy, H. H. Gallion, F. R. Ueland, E. J. Pavlik, and R. J. Kryscio. 'The Efficacy of Transvaginal Sonographic Screening in Asymptomatic Women at Risk for Ovarian Cancer'. *Gynecologic Oncology* 77, no. 3 (June 2000): 350–56. doi:10.1006/gyno.2000.5816.
- 232 Van Nagell, John R., Paul D. DePriest, Frederick R. Ueland, Christopher P. DeSimone, Amy L. Cooper, J. Matt McDonald, Edward J. Pavlik, and Richard J. Kryscio. 'Ovarian Cancer Screening with Annual Transvaginal Sonography'. *Cancer* 109, no. 9 (1 May 2007): 1887–96. doi:10.1002/cncr.22594.
- 233 Ludovisi, M., D. Mavrelou, and D. Jurkovic. 'Re: Factors Affecting Visualization of Postmenopausal Ovaries: Descriptive Study from the Multicenter United Kingdom Collaborative Trial of Ovarian Cancer Screening (UKCTOCS)'. *Ultrasound in Obstetrics & Gynecology* 43, no. 5 (May 2014). doi:10.1002/uog.13321.
- 234 Gollub, E. L., C. Westhoff, and I. E. Timor-Tritsch. 'Detection of Ovaries by Transvaginal Sonography in Postmenopausal Women'. *Ultrasound in Obstetrics and Gynecology* 3, no. 6 (1 November 1993): 422–25. doi:10.1046/j.1469-0705.1993.03060422.x.
- 235 Chan, An-Wen, Jennifer M. Tetzlaff, Douglas G. Altman, Andreas Laupacis, Peter C. Gøtzsche, Karmela Krleža-Jerić, Asbjørn Hróbjartsson, et al. 2013. 'SPIRIT 2013 Statement: Defining Standard Protocol Items for Clinical Trials'. *Annals of Internal Medicine* 158 (3): 200–207. doi:10.7326/0003-4819-158-3-201302050-00583.
- 236 Endorsement – SPIRIT Statement. 2016. Accessed June 19. <http://www.spirit-statement.org/about-spirit/spirit-endorsement/>.
- 237 Chan, An-Wen, Jennifer M. Tetzlaff, Peter C. Gotzsche, Douglas G. Altman, Howard Mann, Jesse A. Berlin, Kay Dickersin, et al. 2013. 'SPIRIT 2013 Explanation and Elaboration: Guidance for Protocols of Clinical Trials'. *Bmj-British Medical Journal* 346 (January): e7586. doi:10.1136/bmj.e7586.
- 238 Quality Governance in the NHS - A Guide for Provider Boards - Publications - GOV.UK. 2015. Accessed October 9. <https://www.gov.uk/government/publications/quality-governance-in-the-nhs-a-guide-for-provider-boards>.

- 239 NHS England Diagnostic Imaging Dataset (DID) 2013-14 [s3.5.5]
- 240 *The Economist*. 2015. 'The Cheap, Convenient Cloud', April 18.
<http://www.economist.com/news/business/21648685-cloud-computing-prices-keep-falling-whole-it-business-will-change-cheap-convenient>.
- 241 OSI Homepage. 2016. Accessed May 7. <http://dicom.nema.org/standard.html>.
- 242 OSI Model. 2016. Wikipedia, the Free Encyclopedia.
https://en.wikipedia.org/w/index.php?title=OSI_model&oldid=718800733.
- 243 Internet Protocol Suite. 2016. Wikipedia, the Free Encyclopedia.
https://en.wikipedia.org/w/index.php?title=Internet_protocol_suite&oldid=718912104.
- 244 Ethernet. 2016. Wikipedia, the Free Encyclopedia.
<https://en.wikipedia.org/w/index.php?title=Ethernet&oldid=717305658>.
- 245 'Overview - SPLWeb'. 2016. Accessed October 20.
<https://www.spl.harvard.edu/pages/Overview>.
- 246 'National Alliance for Medical Image Computing'. 2016. Accessed October 20.
<http://www.na-mic.org/>.
- 247 'NAMIC Downloads' 2016. Accessed October 20. <http://wiki.na-mic.org/Wiki/index.php/Downloads>.
- 247 'Engineering - NAMICWeb'. 2016. Accessed October 20. <http://www.na-mic.org/pages/Engineering>.
- 249 'Documentation/Nightly/Developers/Modules - SlicerWiki'. 2016. Accessed October 20.
<http://wiki.slicer.org/slicerWiki/index.php/Documentation/Nightly/Developers/Modules>.
- 250 'MeVisLab: MeVisLab'. 2016. Accessed October 20. <http://www.mevislab.de/>.
- 251 'Fraunhofer MEVIS'. 2016. Accessed October 20. <http://www.mevis.fraunhofer.de/>.
- 252 '2.2. Development in MeVisLab'. 2016. Accessed October 20.
<http://mevislabdownloads.mevis.de/docs/current/MeVisLab/Resources/Documentation/Publish/SDK/GettingStarted/ch02s02.html>.
- 253 '25.2. Modules (C++) Wizard'. 2016. Accessed October 20.
<http://mevislabdownloads.mevis.de/docs/current/MeVisLab/Resources/Documentation/Publish/SDK/MeVisLabManual/ch25s02.html>.
- 254 'XNAT - Home'. 2016. Accessed October 20. <https://www.xnat.org/>.
- 254 'Neuroinformatics Research Group'. 2016. Accessed October 20. <https://nrg.wustl.edu/>.
- 255 Atlassian. 2016. 'Bitbucket | The Git Solution for Professional Teams'. Bitbucket. Accessed October 20. <https://bitbucket.org/>.
- 256 'Git'. 2016. Accessed October 20. <https://git-scm.com/>.
- 257 'XNAT Documentation: XNAT 1.7 Developer Documentation'. 2016. Accessed October 20. <https://wiki.xnat.org/xnat17/xnat-1-7-developer-documentation>.

References

- 258 'XNAT Marketplace'. 2016. XNAT Marketplace. Accessed October 20. <http://marketplace.xnat.org/>.
- 259 'Developing Java Programs with the JDK'. 2016. Accessed October 20. <https://www.java.com/en/download/faq/develop.xml>.
- 260 'Apache Tomcat® - Welcome!'. 2016. Accessed October 20. <http://tomcat.apache.org/>.
- 261 'PostgreSQL: The World's Most Advanced Open Source Database'. 2016. Accessed October 20. <https://www.postgresql.org/>.
- 262 'XNAT REST API Directory - XNAT 1.6.x Documentation - XNAT Documentation Wiki'. 2016. Accessed October 20. <https://wiki.xnat.org/display/XNAT16/XNAT+REST+API+Directory>.
- 263 'Pipeline Engine - XNAT 1.6.x Documentation - XNAT Documentation Wiki'. 2016. Accessed October 20. <https://wiki.xnat.org/display/XNAT16/Pipeline+Engine>.
- 264 'MRC eMedLab'. 2016. Accessed October 20. <http://www.emedlab.ac.uk/>.
- 265 Babbage, Charles. 1864. *Passages from the Life of a Philosopher*. London: Longman, Green, Longman, Roberts, & Green
- 266 Paul A. Longley, Michael F. Goodchild, David J. Maguire, et Al. 2016. 'Geographical Information Systems: Principles, Techniques, Management and Applications, 2nd Edition, Abridged' John Wiley, Chapter 12 Contribution by H. Veregin
- 267 Cai, Li, and Yangyong Zhu. 2015. 'The Challenges of Data Quality and Data Quality Assessment in the Big Data Era'. *Data Science Journal* 14 (0): 2. doi:10.5334/dsj-2015-002.
- 268 'Data Quality'. 2016. Wikipedia. https://en.wikipedia.org/w/index.php?title=Data_quality&oldid=737361932