

PSYD054

Quality Assurance Procedures in  
Non-Obstetric Diagnostic Ultrasound:  
A Study of the Reliability of Current Methods.

Submitted by Peter George Cantin to the University of Exeter

as a thesis for the degree of

Doctor of Clinical Research

In April 2017.

Volume 1 of 1.

Word Count-39317

Student Number-600060577

This thesis is available for Library use on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: .....

# Contents.

## Table of Contents

|  |                                     |
|--|-------------------------------------|
| <b>Abstract.</b>   | <b>Error! Bookmark not defined.</b> |
| <b>Chapter 1.</b>  | <b>11</b>                           |
| <b>Quality Assurance of Reporting in Non-Obstetric Medical Ultrasound.</b>           | <b>11</b>                           |
| <b>A Systematic Review</b>   | <b>11</b>                           |
| 1.1 Introduction.  | 12                                  |
| 1.2 Aims.  | 13                                  |
| 1.3 Methodology.   | 14                                  |
| 1.4 Results.   | 16                                  |
| 1.5 Volume of Literature.  | 16                                  |
| 1.6 Discussion   | 19                                  |
| 1.7 Conclusion.  | 24                                  |
| <b>Chapter 2</b>   | <b>26</b>                           |
| <b>Methodology.</b>  | <b>26</b>                           |
| 2.1. Introduction.   | 27                                  |
| 2.2. Aims and Objectives.  | 28                                  |
| 2.3. Methodology.  | 30                                  |
| 2.4. Ultrasound Image and Report Quality Assessment Tools                            | 34                                  |
| 2.5 Procedure Consort Diagram.   | 38                                  |
| 2.6 Study Procedure.   | 39                                  |
| 2.7. Feasibility.  | 39                                  |
| 2.8 Data Analysis Strategy   | 41                                  |
| 2.9 Sample Size.   | 42                                  |
| 2.10 Contribution to knowledge.  | 42                                  |
| 2.11 Dissemination of Results  | 43                                  |
| 2.12 Ethical Considerations.   | 43                                  |
| 2.13 Risk Assessment.  | 44                                  |
| <b>Chapter 3</b>   | <b>46</b>                           |
| <b>Rationale for the Study Design.</b>   | <b>46</b>                           |
| 3.1 Introduction   | 47                                  |
| 3.2 Use of a Clinical Outcomes Approach.   | 47                                  |
| 3.3 Retrospective reviewing of Ultrasound Imaging                                    | 48                                  |
| 3.4 Image Reviewing Conditions   | 49                                  |
| 3.5 Use of external reviewers  | 50                                  |
| 3.6 Developing a suitable platform for external peer review                          | 51                                  |
| 3.7 Choice of rating tools for grading the quality of ultrasound images and reports. | 54                                  |
| 3.8 Cohort of study participants.  | 57                                  |
| 3.9 Web-site development   | 59                                  |
| 3.10 Data Analysis Strategy.   | 63                                  |
| 3.11 Recruitment Challenges.   | 64                                  |
| 3.12 Ethical Considerations.   | 65                                  |

|  |            |
|--|------------|
| <b>Chapter 4</b>   | <b>68</b>  |
| <b>Results: Quantitative Data</b>  | <b>68</b>  |
| 4.1 Results Overview   | 69         |
| 4.2 Reviewer Demographics.   | 69         |
| 4.3 Distribution of Studies Reviewed.  | 73         |
| 4.4 Descriptive Statistics for the Quality Assessment Tools under Investigation.     | 73         |
| 4.5 Assessment of Distribution Curves for Normality.                                 | 74         |
| 4.6 Assessment of Agreement Between Participants.                                    | 76         |
| 4.7 Effect of Participant Characteristics on Quality Scores Given.                   | 78         |
| 4.8 Effect of Individual Factors on Ultrasound Examination Ratings.                  | 79         |
| 4.9 Correlation between Quality Measurement Tools.                                   | 83         |
| <b>Chapter 5</b>   | <b>85</b>  |
| <b>Discussion of Quantitative Results.</b>   | <b>85</b>  |
| 5.1. Introduction.   | 86         |
| 5.2 Inter-rater Reliability of Ultrasound Reviews.                                   | 88         |
| 5.3 Systematic Effects of Participant Characteristics on Quality Scores given.       | 101        |
| 5.4 Is modelling possible in ultrasound quality assurance?                           | 107        |
| 5.5 Who should undertake quality assurance?  | 109        |
| <b>Chapter 6</b>   | <b>110</b> |
| <b>Analysis of Participant Comments.</b>   | <b>110</b> |
| 6.1 Introduction   | 111        |
| 6.2 Analysis of Participant Comments   | 112        |
| 6.3 Analysis of Comments Strategy.   | 112        |
| 6.4 Constructiveness of Comments.  | 113        |
| 6.5 Methodology of Comment Analysis.   | 117        |
| 6.6 Statistical Analysis.  | 118        |
| 6.7 Results.   | 119        |
| 6.8 Constructiveness of Comments.  | 120        |
| 6.9 Differences between specialist and non-specialist assessors in comment analysis. | 124        |
| <b>Chapter 7</b>   | <b>125</b> |
| <b>Discussion of Participant Comments.</b>   | <b>125</b> |
| 7.1 Introduction   | 126        |
| 7.2 Analysis of comments.  | 127        |
| 7.3 Use of anonymous feedback.   | 129        |
| 7.4 Qualitative feedback mechanisms  | 130        |
| 7.5 Application of feedback guidance.  | 134        |
| 7.6 Feedback guidance and study conditions.  | 136        |
| 7.7 Training in Giving Feedback.   | 138        |
| 7.8 Who should provide feedback?   | 139        |
| 7.9 Conclusion   | 142        |
| <b>Chapter 8.</b>  | <b>145</b> |
| <b>Recommendations for Clinical Practice.</b>  | <b>145</b> |
| 8.1 Introduction.  | 146        |
| 8.2 Levels of inter-rater agreement.   | 146        |
| 8.3 Systematic differences between groups.   | 148        |

|   |            |
|---|------------|
| 8.4 Use of the internet as a mechanism of undertaking quality assurance and disseminating feedback.                             | 149        |
| 8.5 Qualitative feedback by reviewers.  | 150        |
| 8.6 Selection of Audit Tools.   | 152        |
| <b>Recommendations for Further Research.</b>  | <b>154</b> |
| 9.1 Introduction.   | 155        |
| 9.2 Use of the internet at a tool to improve clinical outcomes in diagnostic ultrasound.  | 155        |
| 9.3 Benefits of formal training in giving feedback to reviewers of ultrasound imaging.  | 156        |
| 9.4 Detailed modelling of the effects of participant experience on quality assurance scores given.                              | 157        |
| 9.5 Who should undertake quality assurance?   | 157        |
| <b>Chapter 10.</b>  | <b>159</b> |
| <b>Study Limitations and Research Learning Opportunities.</b>   | <b>159</b> |
| 10.1 Introduction   | 160        |
| 10.2 Bias.  | 161        |
| 10.3 Sample size.   | 164        |
| 10.4 Recruitment of study subjects.   | 166        |
| 10.5 Intellectual Property.   | 167        |
| <b>Chapter 11</b>   | <b>170</b> |
| <b>Conclusions and Summary.</b>   | <b>170</b> |
| 11.1 Introduction   | 171        |
| 11.2 Original Study Objectives.   | 171        |
| 11.2.1 To create, develop and evaluate a standard image bank and audit tool to facilitate standardisation in quality assurance. | 171        |
| <b>References</b>   | <b>178</b> |
| <b>Appendices</b>   | <b>190</b> |
| Appendix 1.   | 191        |
| STARD checklist for reporting of studies of diagnostic accuracy   | 191        |
| Appendix 2.   | 192        |
| Data Extraction Sheet.  | 192        |
| NHS Research and Development Approval.  | 193        |
| Appendix 4.   | 194        |
| Caldicott Guardian Approval.  | 194        |
| Appendix 5.   | 195        |
| Participation Invitation Sheet.   | 195        |

# TABLE OF TABLES.

|   |                                     |
|---|-------------------------------------|
| TABLE 1. SUMMARY OF STUDIES INCLUDED IN SYSTEMATIC REVIEW.....  | <b>ERROR! BOOKMARK NOT DEFINED.</b> |
| TABLE 2. IMAGE QUALITY ASSESSMENT TOOL. AQP CONTRACT.....   | 33                                  |
| TABLE 3. REPORT QUALITY ASSESSMENT TOOL. AQP CONTRACT.....  | 34                                  |
| TABLE 4. LOCAL REPORT QUALITY ASSESSMENT TOOL USED IN PRINCIPAL RESEARCHER'S DEPARTMENT.....                              | 36                                  |
| TABLE 4.5. CLINICAL ADVICE QUALITY ASSESSMENT TOOL. AQP CONTRACT'.....  | 37                                  |
| TABLE 5. RISK ASSESSMENT OF RESEARCH PROJECT.....   | 44                                  |
| TABLE 6. LOCAL REPORT RATING TOOL. PRINCIPAL RESEARCHER'S INSTITUTION. ....   | 55                                  |
| TABLE 7.APQ IMAGE QUALITY SCORING TOOL.....   | 55                                  |
| TABLE 8. AQP REPORT QUALITY SCORING TOOL.....   | 56                                  |
| TABLE 9. AQP CLINICAL ADVICE QUALITY SCORING TOOL .....   | 57                                  |
| TABLE 10. SUMMARY OF SCORES FOR EACH QUALITY RATING TOOL FOR 22 ULTRASOUND SCANS. ....                                    | 74                                  |
| TABLE 11. ASSESSMENT OF NORMAL DISTRIBUTION OF SCORES FOR EACH QUALITY RATING TOOL.....                                   | 74                                  |
| TABLE 12. ASSESSMENT OF NORMAL DISTRIBUTION OF SCORES FOR EACH QUALITY RATING TOOL AFTER LOG TRANSFORMATION.....          | 76                                  |
| TABLE 13. STRENGTH OF AGREEMENT IN ICC.....   | 75                                  |
| TABLE 14. AGREEMENT BETWEEN 5 PARTICIPANTS USING DIFFERENT QUALITY RATING TOOLS TO ASSESS 22 ULTRASOUND EXAMINATIONS..... | 77                                  |
| TABLE 15. REGRESSION FOR QUALITY RATING TOOL: OVERALL SCORE.....  | 79                                  |
| TABLE 16. REGRESSION FOR LOCAL REPORT QUALITY RATING TOOL.....  | 79                                  |
| TABLE 16.5. REGRESSION FOR THE AQP REPORT QUALITY TOOL.....   | 80                                  |
| TABLE 17. REGRESSION FOR THE IMAGE QUALITY RATING TOOL.....   | 82                                  |
| TABLE 18. CORRELATION BETWEEN ULTRASOUND QUALITY RATING TOOLS.....  | 83                                  |
| TABLE 19. SUMMARY OF EFFECT OF PARTICIPANT CHARACTERISTICS ON SCORES GIVEN. ....  | 101                                 |
| TABLE 20. ASSESSMENT TOOL; AQP REPORT QUALITY. ....   | 103                                 |
| TABLE 21. LOCAL REPORT QUALITY TOOL. ....   | 103                                 |
| TABLE 22. COMMENTS BY SUBJECT AND LEVEL OF PARTICIPANT EXPERTISE.....   | 120                                 |
| TABLE 23. CONSTRUCTIVENESS OF COMMENTS PER PARTICIPANT EXPERTISE AS JUDGED BY AN ULTRASOUND SPECIALIST. P<0.01.....       | 121                                 |
| TABLE 24. CONSTRUCTIVE VS NON-CONSTRUCTIVE COMMENTS AS JUDGED BY A NON-ULTRASOUND SPECIALIST. P=0.48 .....                | <b>ERROR! BOOKMARK NOT DEFINED.</b> |

# TABLE OF FIGURES.

|  |     |
|--|-----|
| FIGURE 1. VISUAL ANALOGUE ASSESSMENT TOOL.....   | 32  |
| FIGURE 2. FLOWCHART SUMMARISING STEPS OF STUDY. ....   | 38  |
| FIGURE 3. VISUAL ANALOGUE ASSESSMENT TOOL.....   | 57  |
| FIGURE 4. SCREENSHOT OF ULTRASOUND REVIEW WEBPAGE.....   | 62  |
| FIGURE 5. NUMBER OF REVIEWS UNDERTAKEN BY EACH PROFESSION. ....  | 70  |
| FIGURE 6. NUMBER OF REVIEWS PER CLINICAL BAND (SONOGRAPHERS ONLY).....   | 71  |
| FIGURE 7. CLINICAL GRADE (AFC) OF EXPERT AND PEER GROUP REVIEWERS.....   | 72  |
| FIGURE 8. NUMBERS OF REVIEWS UNDERTAKEN BY EXPERT AND PEER STUDY PARTICIPANTS. ....                                    | 71  |
| FIGURE 9. NUMBER OF REVIEWS COMPLETED BY EACH STUDY PARTICIPANT.....   | 73  |
| FIGURE 10. CONSTRUCTIVE COMMENT FOR HIGH QUALITY ULTRASOUND EXAMINATION. ....  | 115 |
| FIGURE 11. CONSTRUCTIVE FEEDBACK FOR ULTRASOUND EXAMINATION OF LOWER QUALITY.....                                      | 115 |
| FIGURE 12. COMMENT JUDGED AS NON-CONSTRUCTIVE BY BOTH ASSESSORS.....   | 116 |
| FIGURE 13. NON-CONSTRUCTIVE FEEDBACK, RE-REPORTING THE ULTRASOUND EXAMINATION. ....                                    | 116 |
| FIGURE 14. COMMENT CONTENT FOR ALL PARTICIPANTS. ....  | 119 |
| FIGURE 15. CONSTRUCTIVENESS OF COMMENTS AMONG STUDY PARTICIPANTS.....  | 121 |
| FIGURE 16. CONSTRUCTIVENESS OF COMMENTS AS JUDGED BY A NON-SPECIALIST OBSERVER. ....                                   | 122 |
| FIGURE 17. COMPARISON OF SPECIALIST AND NON-SPECIALIST COMMENTS ANALYSIS.....  | 124 |
| FIGURE 18. NEGATIVE COMMENTS MADE BY A STUDY PARTICIPANT. ....   | 135 |
| FIGURE 19. GRAPH DEMONSTRATING FREQUENCY DISTRIBUTION OF NUMBER OF ULTRASOUND<br>EXAMINATIONS REVIEWED PER STUDY ..... | 166 |

## LIST OF ABBREVIATIONS

|       |   |
|-------|---|
| AQP   | Any Qualified Provider                                    |
| BMUS  | British Medical Ultrasound Society                        |
| CASE  | Consortium for Accreditation of Sonographic Education     |
| GMC   | General Medical Council                                   |
| GP    | General Practitioner                                      |
| ICC   | Intra-class Correlation Coefficient                       |
| NOUS  | Non-Obstetric Ultrasound                                  |
| PACS  | Picture Archiving and Communication System                |
| PC    | Personal Computer   |
| STARD | Standards for Reporting of Studies of Diagnostic Accuracy |
| SWUG  | South-West Ultrasound Users Group.                        |

## **Abstract.**

### **Introduction**

This study aimed to address some of the issues and inconsistencies around clinical quality assurance mechanisms in (non-obstetric) diagnostic medical ultrasound. Quality assurance and resultant quality improvement in this field is sporadic with a plethora of different methodologies, techniques and quality assurance measurements tools. The evidence-base upon which programs are designed is weak with little high quality primary research in this subject area. This study aimed to clarify some of the uncertainties around clinical quality assurance mechanisms in this field of medical imaging.

### **Methodology**

A website was created which allowed the retrospective review of ultrasound imaging and clinical reports to be undertaken on-line. Clinical ultrasound cases were selected which covered a wide spectrum of clinical quality and these cases were uploaded onto this site. Study participants were ultrasound practitioners invited from several professional backgrounds and levels of clinical experience who reviewed this imaging and scored the ultrasound examinations using several different quality assessment tools.

The on-line method of image dissemination facilitated a geographically diverse group of ultrasound practitioners to evaluate the same imaging and clinical reports using the quality assurance measurement tools provided. Outcome measurements included degree of inter-rater agreement between participants for each quality

assessment tool. Systematic differences between different reviewers were also assessed.

Participants were given the opportunity to leave comments regarding the imaging that they had reviewed on the website if they wished. The content and tone of these comments was also analysed.

## **Results**

The inter-rater agreement was classed as 'fair' for all the quality assurance tools under investigation. There was no significant difference between any of the quality measurement tools in terms of inter-rater agreement. Correlation between tools was good.

There were weak systematic differences found between reviewers. Practitioners of more clinical experience rated image quality more highly than those of lesser clinical experience. Practitioners of lower clinical grade tended to rate the quality of clinical report more highly than those of a higher clinical grade.

Participant comments were evenly divided between comments on clinical technique and comments on the quality of the written report. An ultrasound specialist judged that 'expert-group' participants were more likely to give constructive comments than a 'peer-group' of reviewers, but this finding was not confirmed when the comments were analysed by a non-specialist in clinical ultrasound. Overall, there were slightly more constructive comments than non-constructive, but a large proportion of the comments were judged to be non-constructive in nature.

## **Conclusions**

The study demonstrated significant inter-rater variation in quality assessment of diagnostic ultrasound which is probably inherent within the imaging modality itself. Efforts should be directed to managing this variation rather than attempting to eradicate it.

There are some systematic differences between study participants but there was insufficient data to accurately model the precise systematic effects of different participant characteristics and this requires further research with a larger cohort of study participants.

There is scope to improve the quality of feedback to ultrasound practitioners, particularly when this is subjective in nature to maximise the probability of this resulting in positive subsequent change. Formal tuition in the theory and practice of giving feedback should be available to all staff undertaking quality assurance work, irrespective of their degree of expertise, clinical grade or clinical experience.

## **Recommendations for Clinical Practice.**

The following clinical recommendations have been made, based on the evidence gained from this study;

- Quality assurance of non-obstetric ultrasound examinations should ideally be undertaken by those of a higher clinical grade than those undertaking the work being appraised. Peer audit may be acceptable providing there is effective oversight by a senior clinical practitioner.
- The use of a single, expert reviewer in diagnostic ultrasound does not provide adequate assurance in terms of inter-rater reliability and therefore should not

be used. A quality assurance program based on retrospective assessment of ultrasound imaging and clinical reporting should be undertaken by several reviewers to buffer against the effects of inter-reviewer variation.

- Use of the internet provides great advantages in terms of overcoming logistical difficulties in undertaking quality assurance in ultrasound, particularly when undertaken by an external reviewer. However, individual feedback should be given to ultrasound practitioners face-to-face by the senior practitioner responsible for the quality assurance program.
- Those undertaking quality assurance work, regardless of clinical grade and expertise should receive formal training in giving feedback in a constructive fashion. The purpose is to maximise the potential for this feedback to lead to improved clinical standards and outcomes for patients.
- There is currently no primary research evidence to favour one quality assurance tool over another. Quality assurance tool selection may therefore be done at an individual ultrasound unit level. In selection of a quality assurance tool, ultrasound units should consider which tool best reflects the individual requirements and workload of that unit.

# **Chapter 1.**

## **Quality Assurance of Reporting in Non-Obstetric Medical Ultrasound. A Systematic Review**

## **Chapter 1.**

### **Quality Assurance of Reporting in Non-Obstetric Medical Ultrasound.**

#### **A Systematic Review**

**This chapter describes the existing literature around clinical quality assurance in diagnostic ultrasound. It serves to highlight where gaps and flaws within the literature require further research.**

#### **1.1 Introduction.**

Diagnostic non-obstetric ultrasound has historically been the preserve of radiologists, but is now utilised by a wide variety of health-care professionals. Increasing volume and complexity of clinical workload amongst radiologists, along with the need to improve radiology services for patients and clinicians have been powerful drivers in the development of skills-mix in medical ultrasound[1]. In particular, sonographers (radiographers who have undertaken further education and training in ultrasound), now undertake and report medical ultrasound examinations in most UK hospitals[2]. The Royal College of Radiologists has traditionally viewed the independent reporting of medical ultrasound examinations by sonographers as a delegated medical task[3] and this practice therefore falls under guidance issued by the General Medical Council (GMC) on appropriate delegation[4]. However, later guidance from the Society and College of Radiographers[5] makes no such distinction between who should produce the clinical report, whether radiologist, reporting radiographer or other health-care professional providing that the report is of a satisfactory standard. Nonetheless, the issue of quality assurance for sonographers is important in order to guarantee safe and effective practice to service users. In the current political

context, demonstration of robust quality assurance methods will be important to ensure the viability of current ultrasound providers[6, 7].

Quality assurance of ultrasound scans and reports is difficult to achieve. Ultrasound is viewed as safe and easily available[8], and is consequently often used as a first-line investigation in patients with vague and non-specific symptomology, particularly from primary care[9]. This makes the selection of appropriate and robust outcome measures difficult for studies of quality assurance, as negative studies will often have no further investigations or treatment, by which a direct comparison can be made. Moreover, ultrasound is renowned for its operator dependence[10]. It is a 'real-time' imaging modality, relying more heavily on immediate interpretation of the moving ultrasound image than later review of static 'hard-copy' imaging. While retrospective analysis of hardcopy imaging is a long-established and generally effective method of assessing report accuracy for many medical imaging modalities[11-14], this approach should be used with caution in medical ultrasound, given the difficulties inherent in retrospectively reviewing static ultrasound imaging.

This systematic review seeks to explore more deeply the issues surrounding quality assurance of ultrasound studies performed by sonographers and non-radiologists.

## **1.2 Aims.**

The aim of this literature review was to identify and appraise the current knowledge base around quality assurance of medical ultrasound examinations performed by sonographers and other health-care professionals in the non-obstetric setting. The following aims of the systematic review were proposed;

- To assess the scope, age and quality of literature around quality assurance of non-obstetric ultrasound.

- To identify the specific criteria used to judge the quality of sonographer performed ultrasound examinations.
- To find out what reference standards are used to judge these criteria
- To identify likely biases within the literature and suggest appropriate further research methodologies to address these biases.

### **1.3 Methodology.**

#### **1.3.1 Literature Search Strategy.**

Suitable literature was identified by searching of electronic databases; Medline, BNI, EMBASE and CINAHL. Hand searching of relevant paper journals was also undertaken. The reference lists from acquired relevant papers were scrutinised with follow-up of relevant references not identified by other search methods.

Electronic databases were searched using the following search terms and Boolean operators;

(ultraso\* OR sonograph\*) AND reporting AND (accuracy OR audit OR error). The 'explode' function was utilised to widen the scope of the search terms to other potentially useful MeSH headings. Search criteria specified only papers in English. No restrictions were made on the age of relevant papers to ensure that the scope of the review was as wide as possible.

Preliminary screening of the electronic search results was by review of publication titles. Electronic abstracts were retrieved for those titles potentially meeting the inclusion criteria. Full paper or electronic copies of potentially relevant papers were retrieved after secondary screening of electronic abstracts. A final screen of full

papers was then performed and the papers meeting the inclusion criteria were used for data extraction.

Criteria for inclusion were all the following;

- 1) Studies assessing reporting quality in non-obstetric ultrasound;
- 2) Studies performed by sonographers or non-radiological medical practitioners;
- 3) A clear and explicit reference standard.

### 1.3.2 Assessment of Study Quality

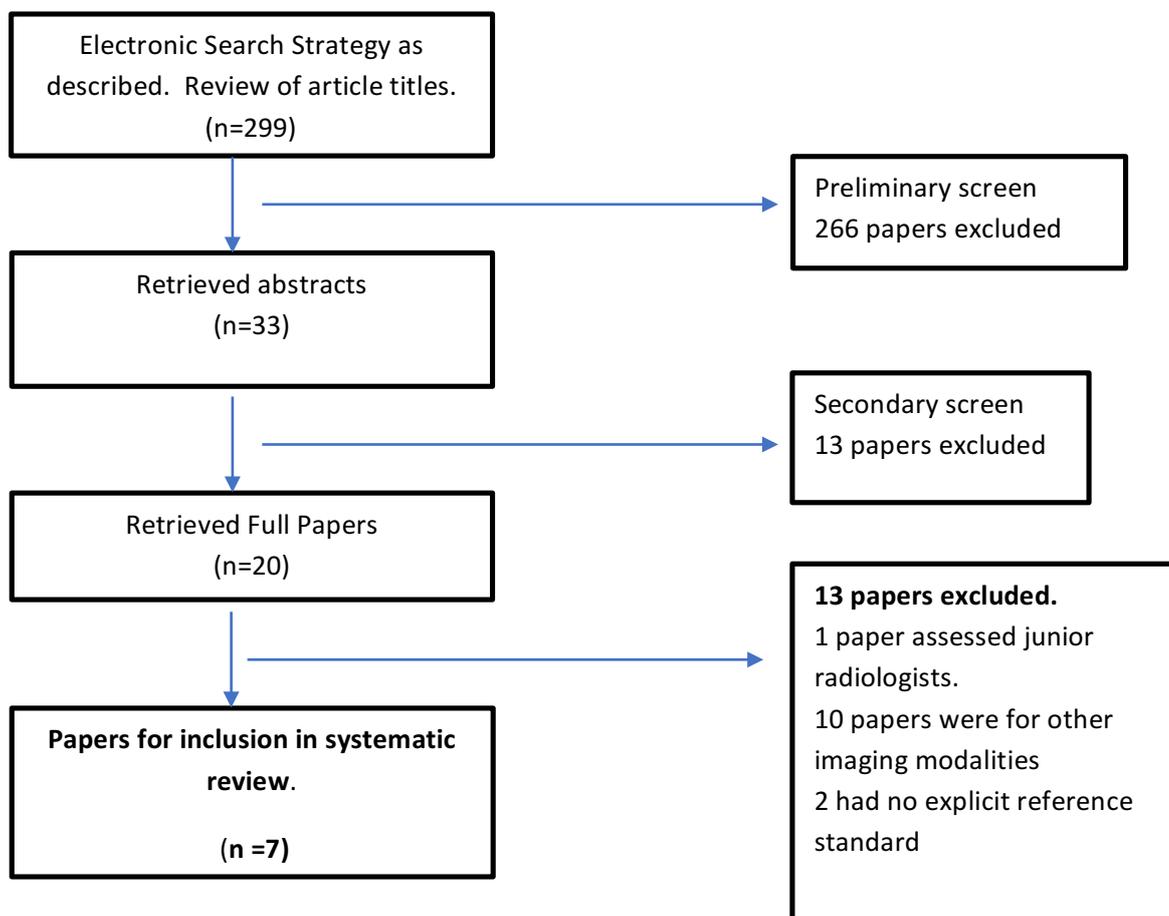
The quality of reviewed papers was assessed with reference to the STARD criteria[15] (Standards for Reporting of Studies of Diagnostic Accuracy). This standard was chosen as it is a well-recognised and validated tool which is explicitly designed to review the quality of studies of diagnostic accuracy. It has clear, unambiguous criteria against which such studies can be assessed. The STARD checklist comprises 25 criteria (appendix 1) against which the quality of studies in diagnostic efficacy can be established. All included studies were judged against the STARD criteria and given a score out of 25. All criteria were evenly weighted.

### 1.3.3 Data Extraction

A data extraction sheet was designed to allow for methodical and consistent extraction of data from all studies included within this literature review. This included extracting details around the reference standard, numbers of participants, outcome measures, and participation criteria. Whether the study was prospective or retrospective was also recorded. All data screening, extraction and assessment of quality was performed by a single person (author). The data extraction sheet is given in appendix 2.

## 1.4 Results.

The search results are summarised below.



**Results for literature review for Quality Assurance in Non-Obstetric Ultrasound**

## 1.5 Volume of Literature.

The volume of literature regarding quality of sonographer clinical reporting was small. Most papers (266) were excluded after preliminary screening of the article titles. The reason for such a large number of excluded papers was that the MeSH headings were deliberately left quite wide to ensure that the maximum number of potentially relevant studies could be included within the literature review. This did result in many studies which did not meet the inclusion criteria on preliminary screening of article titles, hence the large number of studies excluded on preliminary

screening. Even utilising such wide inclusion criteria, only seven studies were identified which met the inclusion criteria. Publication dates ranged from 1994-2010. A summary of the reviewed papers is shown in **Error! Reference source not found.**

- 1.

**Table 1.**

**Summary of Studies Included in Systematic Review.**

| <u>Study</u>                   | <u>Anatomical area</u>                            | <u>Reference standard</u>   | <u>Outcome Measure</u>                          | <u>No of scans and results</u>  | <u>Prospective ?</u> | <u>Quality score (n/25)</u> |
|--------------------------------|---|---|---|---|----------------------|-----------------------------|
| <b>Bates et al 1994[16]</b>    | Abdominal, gynaecological, Superficial structures | Radiologist review of imaging. Review of patient notes  | Accuracy of report                              | n=1046. 94% accuracy  | R                    | 9                           |
| <b>Reynolds et al 2001[17]</b> | Neonatal cranial ultrasound                       | Independent panel of experts  | Accuracy in reporting and prognostic indicators | 42 practitioners reviewed the same scans. 59% of scans correctly interpreted  | P                    | 18                          |
| <b>Weston et al 1994[18]</b>   | Abdominal, gynaecological, superficial structures | Patients scanned by both radiologist and sonographer. Comparison of findings.                                     | Accuracy in reporting                           | n=100<br><br>Overall 20% discrepancy rates. 12% sonographers; 8% radiologists | P                    | 11                          |
| <b>Riley et al 2010[19]</b>    | Musculo-skeletal ultrasound                       | Consultant Radiologist review of images   | Accuracy  | N=248<br><br>Accuracy of 94.8%  | R                    | 10                          |
| <b>Leslie et al 2000[20]</b>   | Abdominal ultrasound                              | All patients scanned by radiologist and sonographer. Comparison of results  | Agreement and accuracy                          | n=100<br><br>Kappa=0.88<br><br>Accuracy -96%                                  | P                    | 13                          |
| <b>Dongola et al 2003[21]</b>  | Abdominal ultrasound                              | Review of notes and correlation with other investigations where possible. Consultant Radiologist review of images | Accuracy  | 104 patients.<br><br>Accuracy of 90.4%  | R                    | 16                          |
| <b>Bude et al 2006[22]</b>     | Abdominal Ultrasound                              | Radiologist review of imaging   | Accuracy  | 5683<br><br>Accuracy of 96.1%   | R                    | 9                           |

Table 1. Summary of studies included in systematic review.

## **1.6 Discussion**

### **1.6.1 Quality Assessment of Reviewed Papers.**

When judged against the STARD criteria, the overall standard of reviewed papers is poor. (Mean score=12/25. Range=9-18). Four of the seven studies were retrospective department audits of sonographer performance[16, 21, 23, 24]. These papers lacked the rigour of primary research papers, which may explain the poor quality of some of these papers when judged against STARD criteria. Common issues included lack of information on participant selection and methodology and poor reporting of results, in particular, reporting of confidence intervals which were reported by only one study[20]. The number and type of ultrasound abnormalities were generally well described, but disease severity was not specifically addressed. This could be inferred from some studies, as some papers reported whether patients were out-patients or in-patients (one would expect to see more severe disease in those patients admitted to hospital), yet a quantitative assessment of disease severity is not possible from this literature.

### **1.6.2 Heterogeneity of Literature.**

The lack of reported confidence intervals makes pooling of data difficult. There is some heterogeneity in the reference standard. The most common reference standard is expert review of scans or hard copy imaging but two studies also incorporated use of medical notes and results of further investigations to assess accuracy of reports. One study used an independent panel of experts to assess the imaging prior to the study. There was some variation in the anatomical area assessed. Five studies assessed 'general medical' ultrasound; i.e. ultrasound of abdominal, gynaecological and superficial structures. One study looked exclusively

at ultrasound of the neonatal brain and one study was concerned with musculo-skeletal ultrasound. The level of heterogeneity within the reviewed literature is therefore significant, making direct comparison of results between papers difficult.

### 1.6.3 Criteria used to Assess Accuracy of Sonographer Reporting.

All the studies sought to assess the accuracy of sonographer reporting as part of delegated radiological care. Most papers (n=6) attempted to introduce a grading system to grade the severity of discrepancies. This ranged from a binary system[18] (significant/not significant) to a 4 part Likert grading system[23]. However, there was insufficient explanation as to how such grading criteria were applied.

It is important to recognise that diagnostic accuracy is not the only indicator of quality of medical imaging reports; wider factors of reporting style are also important. As well as being diagnostically accurate, factors such as content, readability and relevance of the radiology report are now seen as being increasingly important[25-29]. None of the papers within this review addressed the wider quality aspects of ultrasound reporting, being concerned only with diagnostic accuracy. This may reflect the age of some of the papers, with concern over report quality (rather than simple accuracy) appearing to be a relatively new phenomenon. Most of the literature within this review suggests that medical ultrasound undertaken by sonographers is diagnostically accurate. However, there is little literature that evaluates sonographer reporting style and how this may affect the way in which an ultrasound report is interpreted by the referring clinician.

One paper (not included in this systematic review) did attempt to address these issues[30]. This suggests that while sonographers do indeed produce reports which are diagnostically accurate, there is considerable disparity in the way that

sonographers and radiologists report ultrasound studies. The evidence is that sonographers place more 'caveats' or 'disclaimers' within their reports, and do not always provide a clear negative or positive diagnosis to the clinical question. This raises concern that the outcome measures used in studies within this systematic review may be incomplete. It is possible that sonographers are producing reports which are diagnostically accurate, but descriptive, rather than interpretive in nature. While these will score highly on the chosen outcome measures, this does not necessarily mean that ultrasound reports are of high-quality to referring clinicians. This is an area where further research may be warranted.

#### 1.6.5 Reference Standards

One of the criteria for inclusion in this systematic review was a clear reference standard. In six of the reviewed papers, the reference standard was the opinion of an expert reviewer (radiologist) from the same institution. Two of these six papers also attempted to use other criteria, utilising patient outcome (as documented in the patient record) and/or results of further investigations where done.

While such attempts to widen the reference standards are commendable, this may be a source of bias. Because of ultrasound's perceived safety and availability compared with other imaging modalities[8], it is often performed as a first-line investigation for patients in whom symptoms are often vague and non-specific. Ultrasound is therefore frequently used as a 'rule-out', rather than 'rule-in' test. Ultrasound examinations with negative findings often have no further diagnostic investigations or treatments with which to confirm or refute the ultrasound findings. Utilising other diagnostic tests as a reference standard runs the risk of biasing a

study to those patients with positive ultrasound findings which are subsequently confirmed or refuted with additional tests or treatments.

The practice of comparing ultrasound findings with those of an expert reviewer is understandable, but this practice should be approached with caution. Two studies in this review[21, 25] compared sonographers' and radiologists' findings after both had scanned the patient. While both studies gave slightly different figures for accuracy (96% and 88%), the discrepancies were almost equally divided between radiologist and sonographer. This implies that the error rate for radiologists is almost equal to that of sonographers, throwing some doubt on the validity of utilising expert opinion for assessment of diagnostic accuracy. In the radiological literature, there is the acceptance that diagnosis based on medical imaging is not an exact science; there is a degree of variance in how different reviewers review diagnostic imaging[11, 31, 32].

Utilising expert opinion to review hard copy imaging may be the only practical way of undertaking quality assurance within an ultrasound department, but the limitations and potential confounders of this practice should be recognised. Six of the studies in this review utilised a reviewing radiologist from the same institution. While there may be pragmatic reasons for doing this, it does raise the question of bias and the question as to what is truly being measured. Without an external and independent point of reference, it is possible that these studies are simply reporting an institution's view of what is acceptable, rather than a true reflection of diagnostic performance by sonographers. It is open to question as to whether the results of these studies are generalisable to other ultrasound departments.

Only one study in this review utilised expert reviewers which were independent to the departments and staff being appraised. Reynolds et al[33] utilised a standard image bank, which they showed to different practitioners. All practitioners reviewed the same images. The diagnostic findings of these images had been previously established by an independent group of expert reviewers from different institutions. Interestingly, this method of quality assurance showed markedly different results than other studies, with only 59% of respondents giving correct interpretations. The explanation could be due to differences between this study and others regarding professional background and training of participants. However, it is interesting (and a little alarming) that the one study that sought to standardise the image set and utilise independent reviewers yielded accuracy results which were markedly worse than for other, more established quality assurance methodology. Utilisation of standardised image banks is used extensively in other diagnostic imaging modalities, most notably breast radiology[34, 35]. This methodology is rarely used in diagnostic ultrasound, but using this method of quality assurance has the potential to increase credibility of clinical governance processes in the ultrasound field.

#### 1.6.6 Sources of Bias.

There is also the danger of publication bias. Six of the seven studies described a process of quality assurance within their departments, and although the results are variable, there are generally high levels of agreement between sonographers and radiologists. These results are encouraging, but it should be remembered that these studies are an appraisal of a department's performance. Departments falling short of the reported accuracy levels may be discouraged from publishing their results for

fear of negative public and professional perception. The generally low quality of the studies utilised in this literature review make use of a funnel plot to assess the possibility of publication bias inappropriate and this has not therefore been attempted.

The description of selection criteria for participants and study methodology is poor. This makes reported study methodologies almost impossible to replicate, but in addition raises the possibility of recruitment and selection bias. How images were reviewed is also poorly reported, but there is evidence in at least some studies that there was absence of a credible study protocol, resulting in haphazard study performance. For example, one study[18] only took place when the reviewing radiologist happened to be available amongst his other duties. The lack of robust study protocols and uncertainty regarding study recruitment make the potential for selection bias almost impossible to account for within this systematic review.

### **1. 7 Conclusion.**

This systematic review was performed to assess the current level of knowledge of quality assurance methods for performance and reporting of ultrasound by sonographers. Despite, a large amount of literature on diagnostic accuracy of radiology reporting in general, there is surprisingly little on the quality assurance of sonographer reported general ultrasound scans, even though this practice is now widespread[2].

The available literature is generally of poor quality, mainly concerned with audit rather than truly robust primary research of the subject area. This has resulted in poorly reported studies which are difficult to replicate with a high potential for bias. Accuracy rates are generally well reported, but confidence intervals have not been calculated making meta-analysis and pooling of data very difficult.

There are significant issues of bias, particularly around the independence of expert reviewers. Only one study utilised truly independent reviewers and accuracy results from this study are significantly lower than that of other studies.

Outcome measures are mainly concerned with diagnostic accuracy, yet only one study defined what this meant. Most studies compared findings with that of a radiologist, but details of the structure and format of the ultrasound report produced is not reported. The literature in this systematic review was solely concerned with diagnostic 'accuracy' but this does not necessarily equate to a high-quality report and assessment of other factors of report quality (readability, content and ability to give a clear diagnosis) has not been undertaken.

Overall, this systematic review demonstrates significant gaps in the literature around quality assurance of a sonographer-led ultrasound service. Issues of defined outcome measures, having a truly independent and robust reference standard and defects in study protocol and design need to be addressed before robust data around the true levels of quality assurance of diagnostic ultrasound studies can be inferred.

## **Chapter 2**

### **Methodology.**

## **Chapter 2.**

### **Methodology.**

**This chapter will outline the rationale, methodology and methods used to conduct this study.**

#### **2.1. Introduction.**

A systematic review has highlighted some current issues and questions around quality assurance in non-obstetric diagnostic ultrasound. There is now a requirement for ultrasound providers to provide credible quality assurance data under new commissioning arrangements. However, the current literature suggests that mechanisms for quality assurance in non-obstetric ultrasound remain unsatisfactory; an issue that needs to be urgently addressed. Issues uncovered through the systematic review included poor research quality of studies, different reference standards and lack of independence of reference standards.

There is no evidence to suggest that bench-marking of quality assurance processes by either standardisation of the image set or appointment of external (independent) reviewers is common-place among ultrasound providers. This lack of standardisation means that there is no external point of reference for quality-control of ultrasound studies, making meaningful benchmarking of quality assurance practices virtually impossible.

Other imaging modalities have gone some way to implementing standardised quality assurance procedures. The breast-screening program has very well developed procedures in place for bench-marking standards of clinical practice. This includes utilisation of a standardised image bank which allows all clinical staff to review and

judge their performance against their peers. Information Technology is used to ensure that the *same* radiological images are disseminated and reviewed by all practitioners.

While there are substantial differences between ultrasound imaging and breast radiology, there is no reason why at least some of the principles of quality assurance in breast screening could not be applied to ultrasound.

This proposal therefore sought to pilot and evaluate new methods of quality assurance in diagnostic medical ultrasound. By addressing issues of standardisation of reviewed imaging and independence of reviewing practitioners, it was hoped to provide preliminary information on the feasibility of this approach. A secondary outcome of this study was to assess issues of inter-rater agreement in ultrasound quality assurance and the variables which may affect this.

## **2.2. Aims and Objectives.**

***The aim of this study was to design, pilot and evaluate new methods of undertaking quality assurance of non-obstetric diagnostic ultrasound.***

This was achieved by creation of the following objectives;

- To create, develop and evaluate a standard image bank and audit tools to facilitate standardisation in quality assurance.
- To pilot novel, dynamic IT solutions facilitating distribution and review of such images and enable completion of quality assurance reviews. To explore mechanisms by which quality assurance can be undertaken between departments.

- To determine the degree of inter-rater variation in assessment of ultrasound images and reports. To determine whether there was systematic variation in data based on reviewers' professional background, clinical grade, length of experience, age of equipment within the practitioner's employing institution, and type of review setting in which the ultrasound practitioner worked (acute Trust, teaching hospital, primary care, etc.).
- To compare different audit tools currently in use in the quality assurance of non-obstetric diagnostic ultrasound.

Based on the prior literature review, the following hypotheses were tested.

- There is significant inter-rater variation in the way that different practitioners rate ultrasound images and reports.
- There are significant systematic differences in rating of ultrasound image and report quality due to several independent ultrasound practitioner characteristics. These included; background profession, professional expertise, clinical grade and length of clinical experience.
- Experts in medical ultrasound are more critical of ultrasound reports and images than peer group reviews.
- There is significant correlation between different audit tools when used to rate the same image review set.

## **2.3. Methodology.**

### **2.3.1 Study Design**

The study was primarily quasi-experimental, prospective and cross-sectional in design. It fell within the scope of piloting, feasibility and development stages of research in the MRC complex interventions guidance[36]. It was designed to pilot new methods of undertaking quality assurance and to determine the effect of utilising these methods compared with existing literature. The proposed methodology comprised creating a digitised ultrasound image set which was uploaded onto the internet. Volunteer ultrasound practitioners who agreed to take part as study participants from around the UK analysed and rated the same ultrasound studies using several assessment tools. Outcome measures included; inter-rater variation between reviewers, systematic differences between reviewers based on previously described extraneous variables and comparison of results with published literature. It was hoped that this methodology would help to rectify the methodological weaknesses of current quality assurance practices highlighted in the literature review.

### **2.3.2 Study Setting.**

The study was multi-centre in design. The standard ultrasound image and report sets were taken from the principal researcher's own clinical department with appropriate ethical and Caldicott Guardian approval (appendices 3 and 4). The ultrasound images and accompanying diagnostic report sets were made available to any individual ultrasound practitioner undertaking clinical ultrasound within the UK who were willing to review the images and accompanying reports. Participants invited to review studies included both medical and non-medical practitioners

involved in the performance and reporting of non-obstetric ultrasound. The images were available for review either on-line or using a DVD or thermal paper images. Units invited to participate included acute Trusts, teaching hospitals and primary care (community) units

### 2.3.3 Participants

The target population being studied were health-care professionals undertaking non-obstetric ultrasound. This comprised a large, heterogeneous group comprising radiologists, sonographers, non-medical radiological staff, nurses and educationalists.

A strategy of purposive and snowball sampling was implemented. This included identification and utilisation of networks (radiological, educational, radiographic and nursing) who undertake non-obstetric ultrasound. These included;

- Consortium of Accreditation of Sonographic Education (CASE).
- South West Ultrasound Users Group (SWUG) forum.
- College of Radiographers Ultrasound group.
- College of Radiographers Consultant Group.
- British Medical Ultrasound Society (BMUS) scientific and education committee
- Radiological contacts via contacts with field collaborators.

It was anticipated that by utilising these existing networks for purposive and snowball sampling, a diverse group of participants would take part within the study. This method of participant sampling was undertaken for pragmatic reasons. While it was accepted that snowball sampling had implications for selection bias, given the

piloting and scoping nature of the project, this was considered an appropriate sampling technique.

#### 2.3.4 Inclusion Criteria

Ultrasound practitioners from any background were accepted into the study. This included radiologists, sonographers, medics, education providers and nursing staff. All non-medical ultrasound practitioners had a recognised ultrasound qualification. The pathways for training of doctors are different and therefore the inclusion criteria were different. However, all doctors were in (or had undergone) a recognised medical post-graduation training scheme in radiology or had at least one year's experience of performing ultrasound in their own speciality.

#### 2.3.5 Exclusion Criteria.

- Non-medical practitioners with no recognised ultrasound qualification.
- Medical practitioners with no or little (less than 1 year) direct experience of diagnostic ultrasound.
- Those practitioners who did not give consent to allow their data to be analysed as part of this study.

#### 2.3.6 Sample Size

It was hoped to obtain a sample size of 50 participants. This assumed that around 10 departments would agree to take part. Given that there are 168 acute Trusts in England[37], and an unknown number of community providers, this seemed a conservative, yet realistic estimate.

### 2.3.7 User Consultation.

This was a study of ultrasound reporting quality assurance, so it was not anticipated that patients would be directly consulted around this research project. The British Medical Ultrasound Society is a multi-professional body of ultrasound users, including GPs, radiologists, physicians, radiographers, nurses, educationalists and physicists. There is an established Professional Standards Committee and Scientific and Education Committee comprising groups of recognised national and international experts in diagnostic ultrasound. The principal researcher is a member of both committees and it was thought appropriate that members of these committees be asked to comment on study design as well as to be utilised as an expert group of reviewers with which to benchmark the general results of the project.

### 2.3.8 Data Collection.

Data collection was by way of electronic review of ultrasound images and accompanying reports via the internet, utilising specialist bespoke research software, specially designed for this project. Use of personal computers (PCs) for ultrasound image review purposes is commonplace in the clinical environment and was assessed to give adequate image quality for reviewers. The data collection tool encompassed the following features

- **Collection of demographic information** including region of employment, professional background, length of clinical experience, clinical grade, work setting and information on age of ultrasound equipment in use within a participant's employing institution.

- **Tools to assess the quality of the ultrasound image and report.**

There are several tools which have been proposed in assessing both image and accompanying report quality of diagnostic non-obstetric ultrasound. These are detailed below;

## **2.4. Ultrasound Image and Report Quality Assessment Tools**

### 2.4.1 Assessment of Image Quality.

Table 2 describes an ultrasound image quality evaluation tool which has been implemented as part of commissioning arrangements under the ‘Any Qualified Provider’ scheme[38]. It is used to monitor image quality in non-obstetric diagnostic ultrasound.

| Table 2<br><i>Image Quality Assessment Tool for AQP contracts.</i> |  |
|--|--|
| <b>Score</b>   | <b>Description</b>   |
| 5  | High quality examination   |
| 4  | Reasonable image optimisation but with a few poorer quality images (inappropriate focus, etc.) absent measurements or annotation |
| 3  | Suboptimal images but with evidence that this was due to patient factors and attempts made to address the difficulties           |
| 2  | Poor image quality with inadequate attempts to optimise. Clinical question answered correctly                                    |
| 1  | Poor image quality – unacceptable standard   |

Table 2 Image Quality Assessment Tool. AQP contract [37]

## 2.4.2 Assessment of Ultrasound Written Report Quality.

- **AQP Commissioning Arrangements**

Table 3 shows the tool used to monitor the quality of the written ultrasound report under AQP commissioning arrangements[38]. The clinical report is received by referring clinicians following a (non-obstetric) ultrasound examination.

| Table 3  |   |
|--|---|
| <i>Report Quality Assessment Tool for AQP contracts.</i> |   |
| <b>Score</b>   | <b>Description</b>  |
| 5  | Content and structure optimal   |
| 4  | Essence of report satisfactory – slight modification of emphasis or advice  |
| 3  | Report satisfactory but additional differential diagnosis or advice could have been provided. Unlikely to lead to patient harm  |
| 2  | Discrepancy of measurement or interpretation. No immediate harm to patient but requires amended report  |
| 1  | Unnecessary advice leading to inappropriate further investigation. For example: “can’t exclude malignancy” in clearly defined condition leading to invasive test or one involving ionising radiation when unnecessary. Inappropriate follow up recommended leading to downstream costs and patient anxiety. |
| 0  | Poor report with risk of inappropriate management pathway   |

Table 3. Report Quality Assessment tool. AQP contract [37].

- **Locally used report quality tool (Principal Researcher’s Institution)**

The assessment tool in Table 4 also aims to assess ultrasound quality. It has been utilised in the principal researcher’s employing institution for several years.

| Table 4  |   |
|--|---|
| <i>Local Report Quality Assessment Tool (Principal Researcher's Institution)</i> |   |
| Score  | Description   |
| 5  | Complete agreement with report or only very minor change in wording or focus.         |
| 4  | Report accurate but additional comments required.                                     |
| 3  | Report factually accurate but additional differential diagnoses have not been offered |
| 2  | Disagreement in image interpretation. Report therefore inaccurate                     |
| 1  | Clinical Question not answered or cannot be inferred from report.                     |

Table 4. Local Report Quality Assessment Tool used in Principal Researcher's Department.

- **Visual Analogue Score for Report Quality.**

In addition, a visual analogue scale was used for reviewers to report overall report quality. A previous study[39], showed that visual analogue scales are an effective method of capturing overall quality in ultrasound and this was therefore also used to determine its effectiveness in capturing report quality.

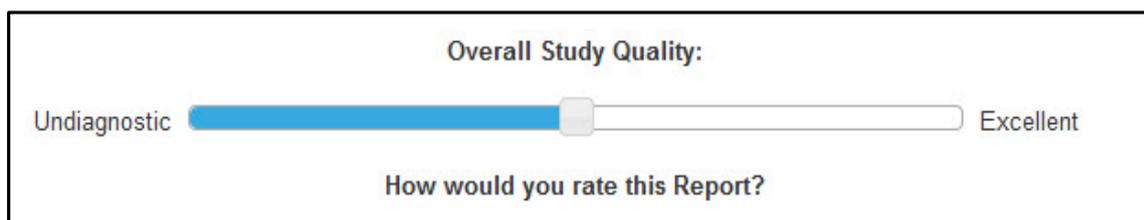


Figure 1. Visual Analogue Tool for Assessment of Diagnostic Ultrasound Studies.

### 2.4.3 Assessment of Quality of Clinical Advice in Ultrasound Report.

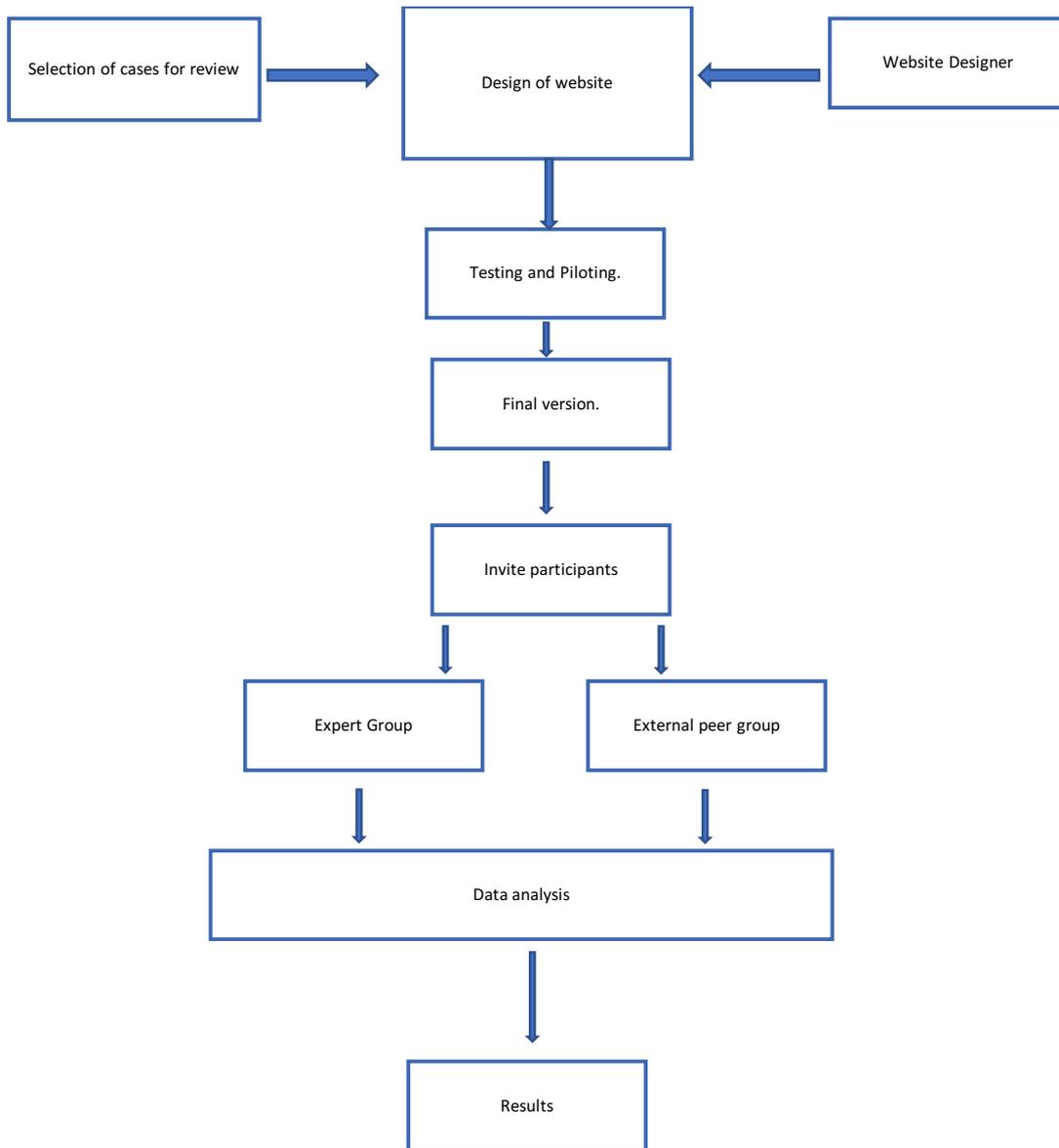
Clinical advice forms an important aspect of the ultrasound report, particularly where the results are equivocal or abnormal. Clinical advice may for example, be a suggestion to undertake further specific diagnostic tests to clarify equivocal findings or to suggest onward referral to a clinical speciality. Clinical advice forms a part of the assessment of ultrasound quality under AQP commissioning arrangements[38] and has therefore been included within this study. The assessment tool is given below in table 4.5.

| Table 4.5  |   |
|--|---|
| <i>Clinical Advice Quality Tool for AQP Contracts.</i> |   |
| <b>Score</b>   | <b>Description</b>  |
| 5  | High quality advice; appropriate further management   |
| 3  | Indeterminate advice. E.g. 'further imaging/investigation recommended.  |
| 1  | Poor advice. Incorrect further management or investigation: e.g. CT when MRI should be advised or CT if no further investigation indicated. |

Table 4.5. Clinical Advice Quality Assessment tool. AQP contract

## 2.5 Procedure Flowchart.

The planned procedure for this study is summarised in the flowchart below. (Figure 2).



---

Figure 2. Flowchart summarising steps of study.

## **2.6 Study Procedure.**

22 diagnostic ultrasound studies were collated. These were studies which have been used previously within the principal researcher's own departmental audit program. The purpose of using images previously used for audit was twofold; a) the risk of incidental detection of serious, unreported pathology was minimised, b) an attempt could be made to distribute studies evenly across a wide spread of ultrasound imaging quality

All patient identifiable information on both images and reports was visibly and electronically removed with specialist software (Clip-Washer; Toshiba, California) and uploaded together with an anonymised ultrasound report to the internet. A web designer was appointed to provide the necessary expertise in web-design to produce a visually attractive and easy to use interface for display of ultrasound imaging and reports. The website was used to display the relevant imaging and to collect individual participant responses to that imaging. The website was hosted by the principal researcher's academic institution (University of Exeter).

Marketing of the survey site was by a combination of advertising within the ultrasound press, via professional groups (RCR, COR, BMUS) and at national conferences. The principal researcher is regularly asked to give presentations at national conferences on quality assurance issues and the study was 'plugged' at these presentations. Local networks (SWUG) were also be used for recruitment. Field collaborators identified radiological input via their own existing networks.

## **2.7. Feasibility.**

It was anticipated that most aspects of this research would be relatively easy to follow-through. Some of the data collection tools have either been in use within a

single department for several years or have undergone piloting in a previous small scale research project. Although undertaking surveys and questionnaires on-line is a well-established research methodology [40-42], this has not yet been used for quality assurance of ultrasound imaging to the principal researcher's knowledge. Use of an expert in website design ensured that the website was fit for purpose.

The principal researcher has a well-established network of contacts among the ultrasound community. The field collaborators have a similarly extensive (but different) network of contacts among radiologists. It was anticipated that this would facilitate adequate study participation with good diversity of study participants.

#### 2.7.1 Reserve Measures.

After risk assessing the project, the aspects of this study thought most at risk of failure were either unexpected IT issues or a lack of study participants. Serious and unresolvable IT issues were thought unlikely with the current sophistication of data collection software. However, the fall-back position was to replicate study images onto DVD and/or thermal paper and distribute images with paper questionnaire copies by post to departments willing to take part. This is the approach used by one study within the literature review<sup>[17]</sup>, with reasonable return rates.

Recruitment difficulties would be addressed by asking contacts within existing networks to 'snowball' knowledge of the study. This is a recognised method of participant recruitment, although it is acknowledged that such techniques can lead to selection bias

## **2.8 Data Analysis Strategy**

Data was stored electronically using an electronic spreadsheet (Excel, Microsoft, California). Data analysis was undertaken using SPSS version 23 (IBM, Texas), a dedicated statistical software package.

Descriptive statistics were applied to participant data such as professional background, length of clinical experience and clinical grade.

### **2.8.1 Assessment of Normality of distribution of Dependent Variables.**

The dependent variables were the scores assigned by study participants to the displayed ultrasound imaging using the review tools under investigation. Skewness and kurtosis were assessed. If data did not show normal distribution, efforts would be made to transform the data to a normal distribution to enable valid parametric statistical testing.

### **2.8.2 Measures of inter-rater agreement.**

As there were multiple reviewers, reviewing multiple ultrasound studies, intra-class Correlation Coefficient was used with 95<sup>th</sup> confidence intervals. This has been successfully used during piloting of outcome measures[39].

### **2.8.3 Effect of independent variables on ultrasound examination ratings.**

Independent variables included professional background, length of clinical experience, clinical grade, external recognition of expertise and age of equipment in the study participant's own clinical institution. Analysis of the effect of these variables in predicting scores for an individual ultrasound examination was performed using linear regression techniques. Significance testing was undertaken with rejection of

the null hypothesis (that the independent variable had no effect on ultrasound study score) at a probability (p) value of 0.05 or less.

#### 2.8.4 Correlation between audit tools.

With the assumption that the data would be normally distributed, correlation between the different audit tools was planned by Spearman's correlation coefficient with rejection of the null hypothesis at a p value of 0.05 or less.

#### **2.9 Sample Size.**

Power analysis was difficult to perform from the existing literature due to difficulties in calculating effect size. Extrapolating data from previous data[39] demonstrated a small-to-medium effect size, estimated at 0.28. Assuming a required significance level of 0.05, estimated sample size was 306 ultrasound scan reviews. It was anticipated that 22 studies would be required to enable fair distribution among each of the 5 report categories (from the principal researcher's own audit tool in use within his own clinical department). This therefore required 14 reviewers to produce results of statistical significance. These calculations were based only on the assessment of image quality rather than report quality. To obtain high quality data on the effect of the described variables, it was hoped to obtain 50 independent reviews of the image set.

#### **2.10 Contribution to knowledge.**

This study is timely in the current political context. Ultrasound providers are now required to provide robust evidence of the clinical quality of their services to

commissioning bodies under the 'any qualified provider' (AQP) scheme. Validated mechanisms for quality assurance of ultrasound studies do not exist at present which is a serious issue for providers and commissioners alike. This study will provide some preliminary pilot work on novel quality assurance mechanisms (such as use of multiple independent reviewers) facilitated by novel IT solutions. It was anticipated that this study would also give high-quality data describing systematic effects of participant variables such as clinical grade, length of clinical experience, and clinical background. This information would be important to model the effects of these systematic factors and build in processes to mitigate their effects.

### **2.11 Dissemination of Results**

Results will be disseminated by presentation at national conference (United Kingdom Radiological Conference and British Medical Ultrasound Society Annual Scientific Meeting). Aspects of this study will be written up for publication in peer-reviewed journals. The choice of journals will reflect different aspects of the study. Scientific data will be submitted to clinical journals such as *Clinical Radiology* or *Radiography* for publication. Mechanisms of quality assurance will be of interest to commissioners, so publication in a journal with a wider readership will also be considered such as *The Health Service Journal*.

### **2.12 Ethical Considerations.**

The study participants were professional staff working within the clinical environment. Preliminary discussions with the researcher's Research and Development Department had indicated the requirement to obtain full ethical approval via the Regional Ethics Committee. One of the potential issues was the use of existing patient diagnostic imaging for research purposes for which explicit patient consent had not been given. All images were anonymised by 'data-scrubbing' of all

patient-identifiable information. In addition, the images were screened by one of the field collaborators to ensure that there were no rare pathologies which could lead to a patient being inadvertently identified by the images. This was discussed with the Regional Ethics Committee Centre manager and the local Trust's Caldicott guardian. This was acceptable providing that the individual undertaking the data-scrubbing would have had access to the imaging in their normal day-to-day duties. Regional Ethics Committee and Caldicott Guardian permission was obtained before commencement of this study (appendices 3,4).

Informed consent for consent was obtained as an opt-in procedure at the start of the web-page that was used for image review. Participants were not able to proceed unless they actively opted-in to the process.

### **2.13 Risk Assessment.**

There were a series of complex steps to be completed before successful conclusion of this project. A risk assessment was therefore undertaken to predict as many risks as possible and attempt to mitigate these to prevent the project being put in jeopardy by unexpected issues. A risk-matrix was developed to define, quantify and mitigate these risks as far as reasonably possibly. This assessment is give in Table 5.

| Table 5   |   |                                       |
|---|---|---------------------------------------|
| <i>Risk Assessment for Research Project.</i>                              |   |                                       |
| <b>Identified Risk</b>  | <b>Management of Risk</b>   | <b>Level of Risk after Mitigation</b> |
| Maintaining confidentiality of patients                                   | Data scrubbing of imaging.<br>Approval of study by Caldicott Guardian and REC.<br>Use of images with common (rather than rare) pathology.<br>Data kept within lockable drawer and on password protected computer.                     | Low                                   |
| Data Loss   | Two copies of data on computer and flash drive (both password protected).   | Low                                   |
| Management and Oversight of Project                                       | Ensure active support from academic and field supervisors.  | Low                                   |
| Feasibility of Project  | Review during development stage by academic supervisor and methodological advisors. Clear and detailed research protocol.<br><br>Clear substitute plans in case of difficulties.  | Medium                                |
| Finances  | Assessment of cost for software for image dissemination and review. Use of undergraduate website designer to keep costs low.<br><br>Awareness of where and how to apply for small grants (e.g. College of Radiographers) if necessary | Medium                                |
| Use of data to assess individual performance of staff at host institution | Appraise data in its entirety. Permanently remove study participants' details during coding to ensure that individual data is not available to any of the research or clinical teams.   | Low                                   |
| Insufficient Time for Completion of Project                               | Regular meetings with academic supervisor and setting of short-term deadlines.<br><br>Ensure protected non-clinical time at work  | Medium                                |
| Poor quality study, not generalisable to wider population                 | Careful and thorough study design. User consultation.<br>Regular meetings with academic and field supervisor.   | Low                                   |

**Table 5. Risk Assessment of Research Project.**

## **Chapter 3**

### **Rationale for the Study Design.**

## **Chapter 3.**

### **Rationale for the Study Design.**

**This chapter is intended to explain the reasons for the methodology chosen for this study. Where compromises have been made, the reasons and justification for these have been made.**

#### **3.1 Introduction**

The literature review has demonstrated several methods of undertaking quality assurance of non-obstetric ultrasound imaging. All methods have advantages and disadvantages but any method utilised in routine clinical practice must be achievable and sustainable as it will need to be undertaken on a regular basis. While the most robust method of quality assurance may be to repeat each ultrasound scan utilising two ultrasound practitioners, there are clear implications in terms of time and resources in using this method, with significantly smaller numbers being examined in the literature<sup>[19]</sup>. There is no published evidence to suggest that any ultrasound department within the UK is using this method for regular quality assurance purposes.

#### **3.2 Use of a Clinical Outcomes Approach.**

Some departments have used a clinical outcomes approach, correlating ultrasound scans with subsequent clinical, biochemical or imaging data to assess the accuracy and efficacy of an ultrasound study and report. While the use of additional, independent clinical outcome data is commendable, this approach does run the risk of biasing a quality assurance program in favour of those patients who have positive ultrasound findings. Patients with negative ultrasound findings with who have no further diagnostic testing cannot realistically be included in such a program.

Because diagnostic ultrasound is viewed as safe and cheap[43], it is frequently used as an initial investigation in patients with vague and non-specific symptomology. Most of these will have negative ultrasound findings, raising serious concerns about the efficacy and comprehensiveness of such an approach. This approach does have its value *in addition to* other quality assurance methods. It can be used to check the accuracy of positive ultrasound findings in the context of a more general quality assurance program and may be used to inform errors meetings or other forms of educational opportunities.

### **3.3 Retrospective reviewing of Ultrasound Imaging**

There is evidence that currently utilised quality assurance methods are utilising electronically displayed 'soft-copy' static imaging in routine quality assurance programs[26]. This has the drawback of attempting to condense a huge amount of real-time information into a series of static images displayed on a computer screen or ultrasound scanner monitor. There is therefore the potential for loss of important information using this method. If an abnormality (present on the real-time image) is not recorded on static imaging, it will not be visible to a subsequent reviewer. Pathological appearances which are missed entirely (rather than correctly or incorrectly interpreted), may not therefore necessarily be recorded within that quality assurance program.

While retrospective reviewing of static ultrasound images clearly has some flaws, it is currently likely to be the only practical method of assessing the quality of ultrasound studies. Modern digital archiving technology does allow the capture and storage of large amounts of real-time (moving) scan information but the time taken for retrospective review of all this information is prohibitive. Recommended ultrasound examination times are 20 minutes per examination for non-obstetric ultrasound[44]

and the recommended sample rate for audit is 5%<sup>[26]</sup>. This would involve an unacceptably high work-load for those undertaking retrospective quality assurance and would be impractical in an everyday, busy clinical environment. It would also involve a large investment in terms of equipment and IT infrastructure to store and display all ultrasound studies in real time.

For this project to be applicable to UK ultrasound departments, quality assurance procedures need to be sustainable on a routine and regular basis. Retrospective electronic review of static imaging at a reviewer's convenience is relatively easy to achieve in clinical practice. It will utilise previously stored ultrasound imaging using existing viewing equipment. There is therefore little need to replace or significantly extend existing IT infrastructure. This method is therefore likely to be the best method of quality assurance, combining practicality and sustainability with data of reasonable quality. This methodology has therefore been chosen in this study.

### **3.4 Image Reviewing Conditions**

Ultrasound images can be reliably and safely reviewed by use of a standard PC and monitor[45]. This has the benefit of enabling ultrasound images and reports to be reviewed in a variety of locations. However, it does make it more difficult to standardise and control for ambient viewing conditions such as background lighting, temperature etc. In this case, a judgement was made that the advantages of allowing study participants to review and undertake the image review remotely, at a time and location convenient to them outweighed the risks in terms of potential variation in ambient viewing conditions. Factors such as participant fatigue are perhaps more important[46] and are outside realistic control of the researchers involved in this study.

### **3.5 Use of external reviewers**

Retrospective reviewing of ultrasound imaging within a single department is relatively straightforward as the infrastructure should be in place to support this within most centres. However, utilising reviewers from the same institution as the scans were performed has the potential to introduce reviewer bias into a quality assurance program. There is no current research evidence to confirm whether this is the case but an internally driven quality assurance program has the potential to simply reveal the collective view of what is acceptable in that particular department[39].

Due to workforce shortages in diagnostic ultrasound, many ultrasound departments are taking an ad-hoc approach to sonographer training with training of new ultrasound practitioners taking place within a single, individual department[47] according to local need. The risk of this is that individual ultrasound departments become increasingly insular and divorced from national influence and professional culture with associated risk of non-recognition of poor practice.

A more critical outside observer may well offer a different standard of acceptable practice in diagnostic ultrasound imaging than an internal observer. Using reviewers external to an organisation does have implications in terms of the sharing of confidential patient information as imaging is not anonymised when stored on conventional Picture Archiving and Communication (PACS) systems. It is also logistically more difficult, with studies having to be 'pushed' and 'pulled' between institutions and between individuals. A mechanism was therefore necessary which facilitated easy external access to anonymised ultrasound images, reports and clinical details. This data would be easily accessible to all study participants at a time and in a place, convenient to them. Displaying images via a website on the internet was the best way of achieving this.

### **3.6 Developing a suitable platform for external peer review**

#### **3.6.1 Types of ultrasound scans appropriate for retrospective review.**

The principal researcher has been undertaking a quality assurance program within his own department since 2008 and the methodology and results of this program have been presented at national level[48]. Experience from this program has helped to identify the type of ultrasound scans suitable for this type of retrospective quality assurance. These include general abdominal, paediatric, gynaecological and superficial-parts ultrasound investigations. Those studies with an integral real-time component such as echocardiography, duplex veno-sonography and musculoskeletal ultrasound are less suitable for retrospective review of static imaging and have been excluded from this study.

#### **3.6.2 Selection of ultrasound scans for review.**

While some would advocate a random selection of studies for review, this was not deemed appropriate for this study. A wide range in quality of ultrasound imaging and reports were felt necessary to produce robust results across all parts of the ultrasound quality spectrum. One of the aims of this study was to determine whether there is agreement between observers about what makes an ultrasound scan good or bad; i.e. is there a universally agreed standard against which to judge diagnostic ultrasound examinations. The principal researcher's experience (in his own institution) with clinical audit demonstrate that the majority of ultrasound studies are of a satisfactory or good standard[48]. Taking a random selection of studies would have therefore introduced bias towards those studies which were of reasonable standard. This data would not be sufficient to determine if different reviewers could consistently differentiate between ultrasound scans of high and low clinical quality.

The ultrasound examinations chosen for this project were therefore carefully selected to enable a wide range of qualities of ultrasound scans to be displayed. This was achieved by reviewing previous audit data from the principal researcher's own institution, ensuring that a wide range of ultrasound examination quality was selected.

### 3.6.3 A means of displaying ultrasound images and accompanying reports, clinical presentation and patient demographics to study participants.

To maximise participation from as many ultrasound practitioners as possible, a means had to be devised which enabled ultrasound studies, patient demographics, clinical histories and ultrasound reports to be accessed quickly and easily and at a time convenient to study participants. Quick, easy data collection tools were also required which combined ease and rapidity of use to facilitate easy participation for study participants.

This was achieved by a decision to display all of this information over the internet. This however, did create a new set of logistical challenges. No such site existed for the distribution of ultrasound images and on-line methods of ultrasound quality data collection. A website therefore needed to be created which would enable display of patient demographics, clinical data, ultrasound imaging and corresponding reports. The site also required the ability to record participant reviews of this data in real-time.

The creation of such a website required collaboration with an individual with sufficient IT expertise to create a website that was visually appealing, easy to navigate and with sufficient ease of use to enhance participant recruitment onto the study. One of the core requirements of this site was functionality that facilitated both

easy review of ultrasound imaging and collation of participant responses on-line and in real-time.

The use of the internet raised some ethical issues around displaying confidential patient data that may be visible to any individual worldwide. There was a challenge in achieving a balance between making the website secure, yet ensuring that study participants had easy access to the site to undertake reviews of the ultrasound imaging.

A potential solution was 'vetting' of individual participants by the principal researcher with creation and distribution of individual passwords for each study participant. However, this would have made the site more difficult to access for potential study participants and increase the time required for such vetting by the principal researcher. Using this technique clearly had negative implications for participant recruitment.

An alternative was to open the site to any potential participant with the attendant risk of data security and inappropriate use by unqualified users. Although this was considered, after exploration with the Research and Development Department and Caldicott Guardian at the principal researcher's employing institution, this was not thought to be appropriate.

The compromise was for individuals to self-register on the site with automatic creation of an individual password for each study participant. Each study participant was required to give some demographic and professional information before being given access to the ultrasound imaging. In addition, all (visible and non-visible) patient-identifiable information was removed by a proprietary software product,

specifically designed for this purpose (Clipwasher, Toshiba, California.). No identifiable patient details were therefore available on-line.

### **3.7 Choice of rating tools for grading the quality of ultrasound images and reports.**

#### **3.7.1 Local Quality Assessment Tool**

There are currently no fully validated tools for the grading of ultrasound images and reports although a wide variety of tools have been proposed[26]. The principal researcher has been undertaking ultrasound quality assurance for several years, using a locally designed and implemented quality assurance tool. This tool was designed to assess clinical report quality alone, with no mechanism for assessing the quality of the ultrasound image. This tool has been in use in the principal researcher's own clinical institution for a period of 8 years and has informed and driven many education activities over this time. While this tool has not been robustly tested to assess inter-reviewer agreement, it was felt important to include this rating tool due to the length of time that it has been in real, clinical use. The tool is shown in Table 6.

| Table 6  |  |
|--|--|
| <i>Local Report Rating Tool. Principal Researcher's Institution.</i> |  |
| <b>Score</b>   | <b>Description</b>   |
| 5  | Complete agreement with report or only very minor change in wording or focus.          |
| 4  | Report accurate but additional comments required.                                      |
| 3  | Report factually accurate but additional differential diagnoses have not been offered. |
| 2  | Disagreement in image interpretation. Report therefore inaccurate.                     |
| 1  | Clinical Question not answered or cannot be inferred from report.                      |

Table 6. Local Report Rating Tool. Principal Researcher's Institution.

### 3.7.2 AQP Contracts Quality Assessment Tool

Quality Assurance in diagnostic ultrasound imaging is an integral part of the 'any qualified provider' (AQP) contracts. This scheme allows any diagnostic ultrasound provider to undertake work for the NHS proving that they meet the criteria stipulated by clinical commissioning groups. Most contracts for providers have a requirement to undertake retrospective quality assurance of clinical quality of service. A quality assessment tool is stipulated together with a scoring system to support its use[38]. This system of quality assurance and the accompanying tool has not been validated yet underpins many AQP contracts.

It is therefore appropriate to test these rating tools within this research project. The AQP contracts require scoring of image quality (table 7), report quality (table 8) and quality of clinical advice (Table 9).

Table 7.

*Image Quality Scoring Tool for AQP Contracts.*

| <b>Score</b> | <b>Description</b>   |
|--------------|--|
| 5            | High quality examination   |
| 4            | Reasonable image optimisation but with a few poorer quality images (inappropriate focus, etc.) absent measurements or annotation |
| 3            | Suboptimal images but with evidence that this was due to patient factors and attempts made to address the difficulties           |
| 2            | Poor image quality with inadequate attempts to optimise. Clinical question answered correctly                                    |
| 1            | Poor image quality – unacceptable standard   |

Table 7. APQ Image Quality scoring tool.

Table 8

*Report Quality Scoring Tool for AQP Contracts.*

| <b>Score</b> | <b>Description</b>  |
|--------------|---|
| 5            | Content and structure optimal   |
| 4            | Essence of report satisfactory – slight modification of emphasis or advice  |
| 3            | Report satisfactory but additional differential diagnosis or advice could have been provided. Unlikely to lead to patient harm  |
| 2            | Discrepancy of measurement or interpretation. No immediate harm to patient but requires amended report  |
| 1            | Unnecessary advice leading to inappropriate further investigation. For example: “can’t exclude malignancy” in clearly defined condition leading to invasive test or one involving ionising radiation when unnecessary. Inappropriate follow up recommended leading to downstream costs and patient anxiety. |
| 0            | Poor report with risk of inappropriate management pathway   |

Table 8. AQP Report Quality Scoring Tool

| Table 9  |   |
|--|---|
| <i>Clinical Advice Scoring Tool for AQP Contracts.</i> |   |
| <b>Score</b>   | <b>Description</b>  |
| 5  | High quality advice; appropriate further management   |
| 3  | Indeterminate advice. E.g. 'further imaging/investigation recommended.  |
| 1  | Poor advice. Incorrect further management or investigation: e.g. CT when MRI should be advised or CT if no further investigation indicated. |

Table 9. AQP Clinical Advice Quality Scoring Tool

### 3.7.3 Continuous Analogue Scoring Tool

There has been some research on methods of assessing image quality in gynaecological ultrasound<sup>[39]</sup>, and this work suggested that a continual analogue scale was the most consistent way of grading ultrasound quality. This was therefore incorporated as a quality assessment tool within the website by means of a sliding bar. Measurement along this bar gave a score from 0-10.

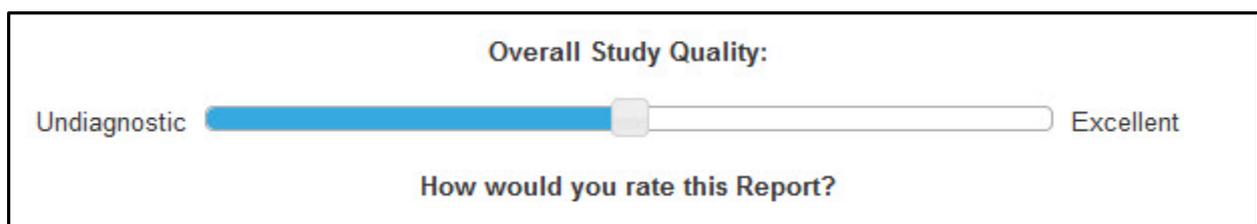


Figure 3. Visual Analogue Assessment Tool

Thus, a wide range of data collection tools were used within the website including categorical, ordinal and continuous data scoring methods.

### 3.8 Cohort of study participants.

The study required that a number of ultrasound practitioners with a range of length of clinical experience, professional backgrounds and seniority (as reflected in clinical

banding) undertake review of several ultrasound examinations and accompanying clinical reports.

During the design phase of the study, participant recruitment was thought to be relatively straightforward, but during the piloting phase it emerged as a significant problem. Although a large amount of time and effort had been invested in designing a website which was as easy as possible for study participants to use, participation within the study still required a significant time commitment (estimated at around two hours by staff piloting the website within my own department). In the current NHS climate, with ultrasound staff being placed under increasing pressure to meet government waiting targets, it was clearly going to be a challenge to incentivise study participants to complete the required reviews.

Methods of clearing this hurdle were discussed with my study supervisor. Even piloting of the website amongst a group of radiography students to check for functionality problems resulted in a poor response rate. Applying for continuing professional development (CPD) credits for participants under the College of Radiographers CPD scheme was considered but reviews by a multidisciplinary audience was required and it was thought that the website would only attract radiographers rather than a truly multidisciplinary cohort of participants which could potentially create bias within the study. This was therefore not done.

The strategy for data collection was therefore changed to ensure that the major study aims could still be achieved, while being achievable in terms of numbers of participants. It was eventually decided to create two specific arms from the cohort of participants;

1. An 'expert-group' drawn from the Professional Standards Committee of the British Medical Ultrasound Society (BMUS). Diagnostic ultrasound is renowned for its operator dependence and subjectivity in interpretation of the ultrasound images, as well as clinical report construction [49]. Trying to implement a reference standard was felt to be important so that the responses from study participants could be compared both with peers and with those from a more expert group. The members of the BMUS Professional Standards Committee are well established ultrasound practitioners, co-opted onto this group because of their recognised and established expertise in diagnostic ultrasound. This was felt to be a suitable group from which to derive an expert assessment of the ultrasound imaging and reports.

2. A 'peer-group' of ultrasound practitioners. This group comprise the main body of current ultrasound practitioners who may well have responsibilities in undertaking audit/quality assurance within their own clinical departments. This group would be expected to undertake clinical work to a similar standard and of a similar clinical complexity to those whose ultrasound examinations were selected for display within this project.

### **3.9 Web-site development**

The website designer was given a brief as to the requirements of the project. To minimise costs of the project, an undergraduate website designer was approached via the IT department of Exeter University. It was agreed that the website design would be undertaken as part of an undergraduate IT student's final year project. The website designer and principal researcher collaborated to

produce a website which was secure yet easy to access, easy to navigate, visually attractive and which displayed static ultrasound images and accompanying ultrasound reports. The quality assessment tools were incorporated within the website and the website had the functionality to collate participant responses on-line and in real-time. The design brief also involved collection of basic demographic data so that systematic effects of the various participant characteristics could be modelled during the data analysis phase of the project.

#### 3.9.1 Technical specifications for the website.

The brief for the website was to produce an easy to use and visually attractive website allowing for the display of patient clinical and demographic data, ultrasound images and accompanying reports together with data collection tools already described. The displayed imaging needed to be large enough that study participants could make a proper assessment of the quality, content and clinical information within each ultrasound image. In addition, a means of retrieving completed reviews through a separate (administrator) area of the website with basic data analysis tools was required.

The website needed to be able to obtain and store participant demographic data that included years of clinical experience, professional background, clinical grade and average age of equipment used in an individual participant's clinical department. In addition, the inclusion of a free-text box was specified to enable study participants to express subjective comments and opinions which could not be easily made using the more quantitative data collection tools.

This was clearly a lot of information to fit onto a web-page and making such a site visually attractive and easy to navigate was a significant challenge. It was not possible to display all the ultrasound images at adequate size on a single webpage yet it was important that study participants could make a brief overall preliminary assessment of the number of ultrasound images as well as the content of those images prior to formal scoring of the ultrasound examination. The solution was to create a series of 'thumbnail' images which allowed study participants to gain an overall preliminary impression of the imaging in its entirety. Clicking on an individual thumbnail image allowed full size display of that image so that it could be properly examined. A screenshot of the webpage (as seen by study participants) is given in Figure 4.

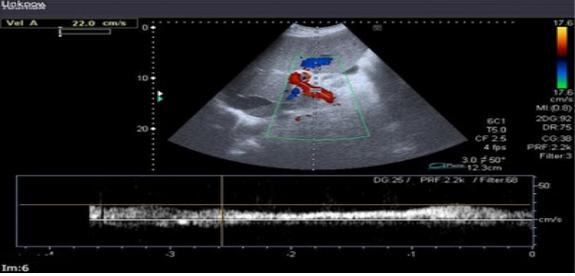
Figure 3. Screenshot of ultrasound review webpage.

Ultrasound image review system

**Abdominal 1**

Please complete the form below to rate this image set. The gallery allows you to view the ultrasound images in detail, and you can download the report using the button below the gallery.



Both kidneys are normal in length. There is mild increased in cortical reflectivity and some loss of parenchymal thickness consistent with the known chronic renal disease.

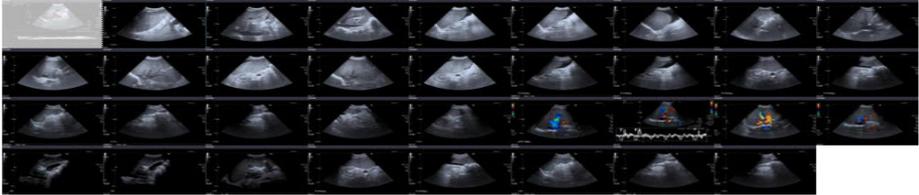
The pancreas was not well visualised due to bowel gas.

No intra-abdominal or pelvic mass lesion identified.

**Conclusion:**

The spleen is enlarged, but this is a non-specific finding. The liver appears normal and hepatic vasculature is patent. No intra-abdominal mass lesion identified.



**Overall Study Quality**

Overall Study Quality:

Undiagnostic 



 Excellent

How would you rate this Report?

- 5 - No change to report.
- 4 - Minor addition to report including comments regarding further management, investigations etc.
- 3 - Additional differential diagnoses not mentioned in the report.
- 2 - Disagreement with interpretation of the ultrasound imaging.
- 1 - Clinical question not answered or cannot be inferred from the report.

How would you rate the image quality?

|   |  |   |   |                                     |
|---|--|---|---|-------------------------------------|
| <b>1 - Poor image quality – unacceptable standard</b> | <b>2 - Poor image quality with inadequate attempts to optimise. Clinical question answered correctly</b> | <b>3 - Suboptimal images but with evidence that this was due to patient factors and attempts made to address the difficulties</b> | <b>4 - Reasonable image optimisation but with a few poorer quality images (inappropriate focus, etc.) absent measurements or annotation</b> | <b>5 - High quality examination</b> |
| <input type="radio"/>                                 | <input type="radio"/>  | <input type="radio"/>   | <input type="radio"/>   | <input type="radio"/>               |

How would you rate the ultrasound report quality?

|  |  |  |  |   |  |
|--|--|--|--|---|--|
| <b>0 - Poor report with risk of inappropriate management pathway</b> | <b>1 - Unnecessary advice leading to inappropriate further investigation. For example: "can't exclude malignancy" in clearly defined condition leading to invasive test or one involving ionising radiation when unnecessary. Inappropriate follow up recommended leading to downstream costs and patient anxiety.</b> | <b>2 - Discrepancy of measurement or interpretation. No immediate harm to patient but requires amended report.</b> | <b>3 - Report satisfactory but additional differential diagnosis or advice could have been provided. Unlikely to lead to patient harm.</b> | <b>4 - Essence of report satisfactory – slight modification of emphasis or advice</b> | <b>5 - Content and structure optimal</b> |
| <input type="radio"/>  | <input type="radio"/>  | <input type="radio"/>  | <input type="radio"/>  | <input type="radio"/>   | <input type="radio"/>                    |

How would you rate the quality of the advice / conclusion in the report?

|   |   |   |                       |
|---|---|---|-----------------------|
| <b>1 - Poor advice – incorrect further management or investigation e.g. CT when MRI should be advised, CT if no further investigation indicated</b> | <b>3 - Indeterminate advice – "further imaging/investigation recommended"</b> | <b>5 - High quality advice – appropriate further management</b> | <b>Not Applicable</b> |
| <input type="radio"/>   | <input type="radio"/>   | <input type="radio"/>   | <input type="radio"/> |

Please enter any additional comments here...

Save Answers

### **3.10 Data Analysis Strategy.**

Data analysis for this study was complex due to the multiplicity of both dependent and independent variables. A statistician was consulted and advice taken. Dedicated statistical software (SPSS v23, IBM, Texas) was used for data analysis.

#### **3.10.1 Use of Intra-class Correlation Coefficient Statistic (ICC)**

Determining levels of agreement between several reviewers, across several ultrasound studies and using several different rating tools was difficult. While Kappa analysis is the most common method of assessing inter-rater agreement[50], this was not appropriate for this study as Kappa analysis cannot be used across multiple reviewers. ICC allows assessment of agreement across multiple reviewers, hence its use within this study.

#### **3.10.2 Use of Regression Techniques.**

This method was chosen to model the systematic effects (if present) of individual participant characteristics on the scores given to individual ultrasound examinations by those study participants. By identifying those factors which have systemic and predictable effects on ultrasound scores and then modelling their precise effect, a model could be created which accounted for these different factors. This would facilitate the use of these audit tools under a range of different conditions and by a range of different ultrasound practitioners.

Analysis of Variance (ANOVA) techniques could be used but this depended upon having a complete data set with no missing values. It was unlikely that every study participant would review all the ultrasound examinations available, due to

the time necessary for undertaking these reviews. Under these circumstances, ANOVA techniques were thought to be unhelpful.

While linear regression techniques overcame this issue, there were problems relating to the categorical (rather than binary) nature of the participant characteristic variables. This was solved by allocating dummy variables to categorical data so that they could be treated as binary data during regression. This is a well-recognised and legitimate technique in overcoming this problem[51].

### **3.11 Recruitment Challenges.**

Recruitment for this project was not initially anticipated to be an issue. A large amount of time and effort had gone into producing a user-friendly website with easy access to ultrasound imaging and data collection tools. Because use of the internet is now widespread among most health care professionals<sup>[52]</sup> during their professional careers, undertaking this study through a website seemed a productive way to undertake this project.

However, piloting had thrown up a significant problem with recruitment. Finding sufficient numbers of individuals who were qualified and willing to spend time reviewing ultrasound imaging was difficult. The completed reviews took approximately 2 hours. The time required for image review proved to be a major hurdle in motivating busy healthcare professionals to undertake the required reviews.

Although the website was advertised at relevant study days and conferences, uptake and recruitment was disappointing. In retrospect, not enough thought and time was put into recruitment for the project, with subsequent requirement for a change in strategy and methodology as the study progressed. Recruitment for clinical studies is a significant hurdle[53], yet there are strategies that could have been better planned to increase recruitment [53-59]. This will be explored further within the discussion section of this project.

### **3.12 Ethical Considerations.**

The study participants were professional staff working within a clinical environment. There were clearly ethical implications in placing confidential radiological imaging and data onto the internet for subsequent review by an unknown number of observers. Permission to undertake this study was obtained from Regional Ethics Committee and the Caldicott Guardian at the principal researcher's employing institution. Specific ethical issues included;

#### **3.12.1 Patient Consent for Use of Imaging.**

The use of confidential patient data in research raises issues of whether consent should be obtained from patients from whom the data was taken. There was some debate as to whether specific consent should be sought from patients on whom the ultrasound imaging had been performed. This was discussed in some detail with the Regional Ethics Committee and Caldicott Guardian. Use of existing data is acceptable without patient consent, providing that the ultrasound images and reports could be robustly unlinked from identifiable patient data by a staff member who would normally access this data[60]. The principal researcher had already accessed this data through his department's own quality assurance

processes. The ultrasound images were 'data-scrubbed' utilising software specially designed for this purpose (ClipWasher, Toshiba, California) to ensure that all traces of confidential data were electronically removed from the imaging before uploading onto the internet. In addition, cases demonstrating rare pathologies were excluded from the data set to reduce the risk of an image set being linked to an individual patient.

### 3.12.2 Participant Consent.

Information was provided to study participants on logging on to the website. The information detailed the purpose of the study and what was expected from study participants. Potential study participants had to opt-into the study before being able to register with the site. Participant information is given in appendix 5.

### 3.12.3 Confidentiality.

All identifiable patient data was unlinked by the principle researcher before uploading onto the internet and 'data-scrubbed' to ensure that no electronic data was embedded within the imaging which could facilitate identification.

Demographic information was obtained from study participants. Demographic data which may have facilitated identification of individual participants (e.g. department worked in) were not mandatory and study participants could therefore choose whether to give potential participant identifying information. Assurances were offered that no individually identifiable data would be utilised within the data analysis or write up aspects of the project.

### 3.12.3 Ethical Considerations in online research.

The use of internet-based research is a relatively new phenomenon which opens some interesting ethical questions. Issues of consent become more problematic as well as knowledge of participants' suitability to take part in such research [55, 61]. However, the distinction does need to be drawn between utilising the internet to access and undertake research within online communities (such as internet forums) and merely utilising the internet as a tool to facilitate easy distribution of images and data collection tool.[61]. The use of the internet as a tool rather than a primary research instrument is therefore a crucial distinction to make. This research is merely using the internet as a tool to distribute and collect data and this use does not therefore have additional ethical implications.

## **Chapter 4**

### **Results: Quantitative Data**

## **Chapter 4.**

### **Results of Quantitative Data.**

**This chapter describes the quantitative results obtained for the ultrasound scans and accompanying reports. This chapter does *not* seek to undertake analysis of free participant comments which will be addressed in a subsequent chapter.**

#### **4.1 Results Overview**

In total, 176 reviews of 22 separate ultrasound studies were completed on the web-based platform. The cases uploaded onto the image review website were as follows: 9 ultrasound studies were of general abdominal ultrasound scans (numbered 1-9 on the website); 8 were gynaecological ultrasound scans (g1-g8) and 5 were ultrasound scans of the urinary tract (u1-u5). The mean length of clinical experience of the reviewers was 21.6 years (SD=7.9years). Maximum length of experience was 39 years and the minimum was 0 years.

One of the rating tools, AQP advice, has not been included within the results. The reason is that many of the ultrasound studies selected for inclusion within this project were straightforward without need for clinical advice to be included within the ultrasound report. The rating tool 'AQP advice' was therefore seldom used by study participants, thus yielding insufficient data for analysis.

#### **4.2 Reviewer Demographics.**

##### **4.2.1 Professional Background.**

Most reviews (91.5%: n=161) were undertaken by non-medical ultrasound practitioners (sonographers). The remainder were undertaken by radiologists (7.9%:

n=14), with only one study being reviewed by an ultrasound educator (0.6%: n=1).

The breakdown of the number of studies reviewed against professional background is given in figure 4.5.

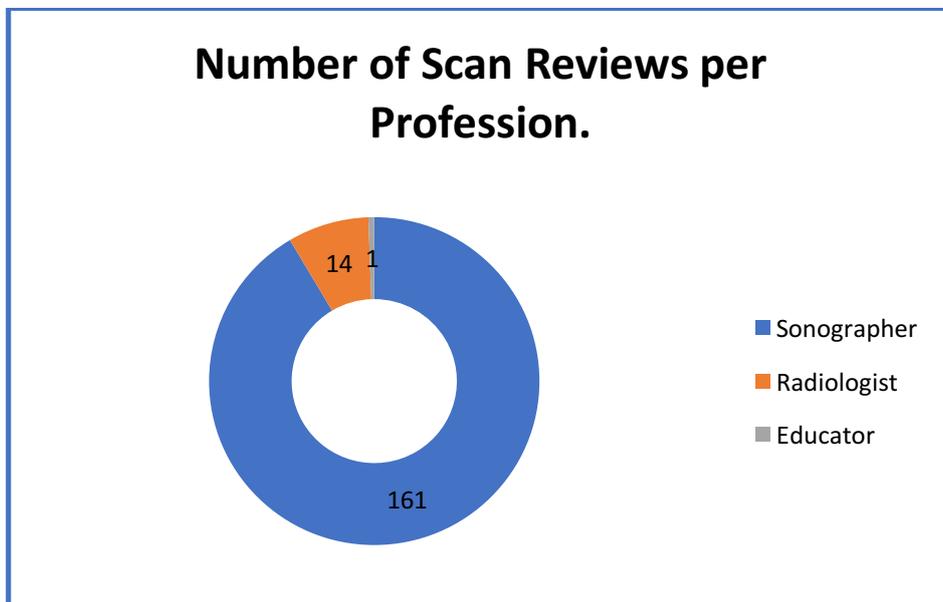


Figure 5. Number of Reviews undertaken by each profession.

#### 4.2.2 Professional Grade.

Four radiologists participated within this study, all of whom were at consultant grade.

Twelve sonographers participated in the project. Of the 161 scan reviews undertaken by sonographers, 78 scans were rated by sonographers at Agenda for Change (AfC) Band 7, 30 scans were reviewed by sonographers at AfC Band 8a and the remaining 53 scans were reviewed by sonographers at AfC Band 8b. This is demonstrated graphically in Figure .

Only 1 ultrasound scan was reviewed by an educator who was at grade 10.

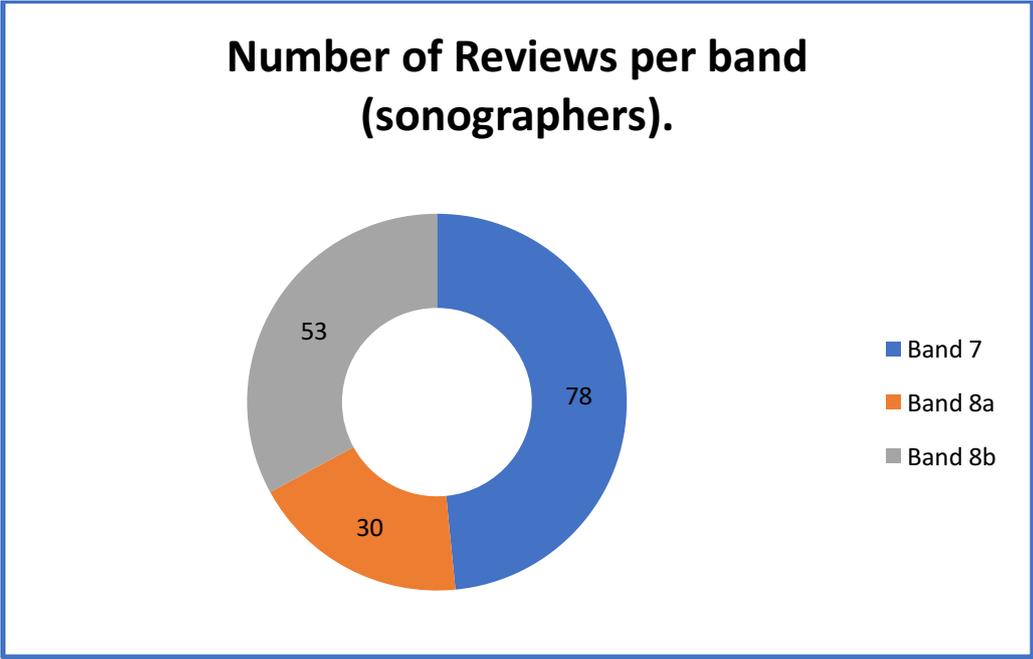


Figure 6. Number of Reviews per Clinical Band (Sonographers only)

4.2.3 Expert and Peer Group Reviewers.

The assumption of expertise was conferred by the following criteria; invited/co-opted membership of the professional standards groups of the British Medical Ultrasound Society, and /or practitioners with significant publication and invited lectureship records at national and international conferences.

Peer group participants were classified as those ultrasound professionals who did not meet these criteria. The professional grade for both the expert and peer groups within this study are presented in Figure :

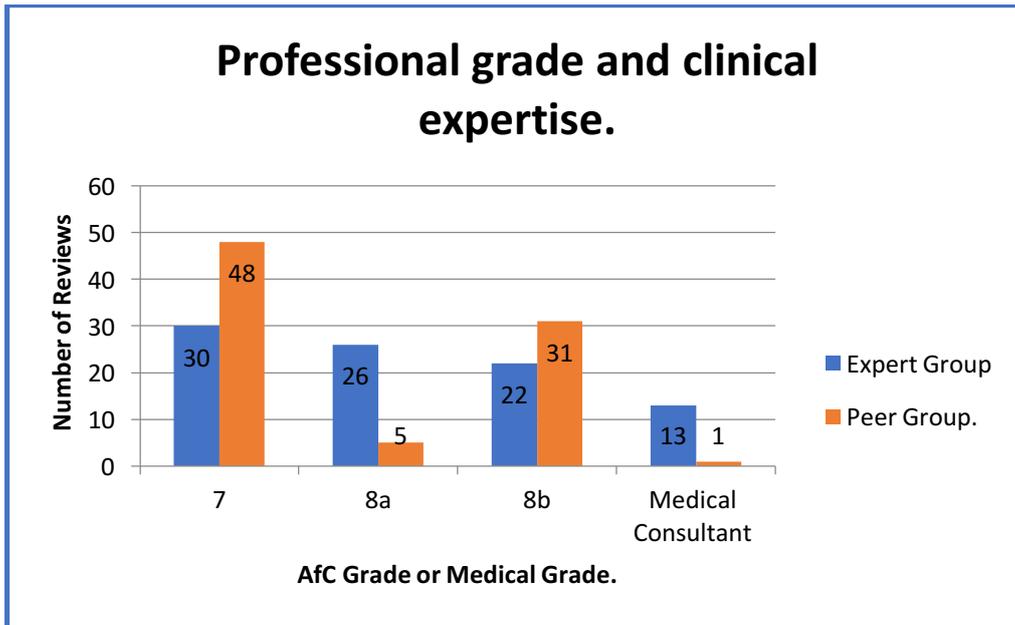


Figure 7. Clinical Grade (AfC) of Expert and Peer Group Reviewers

**4.2.4 Numbers of Ultrasound Examinations reviewed by expert/peer group.**

The numbers of ultrasound examinations reviewed by expert and peer groups were approximately equal with 90 reviews undertaken by the expert group and 86 reviews undertaken by the peer group ( Figure 8).

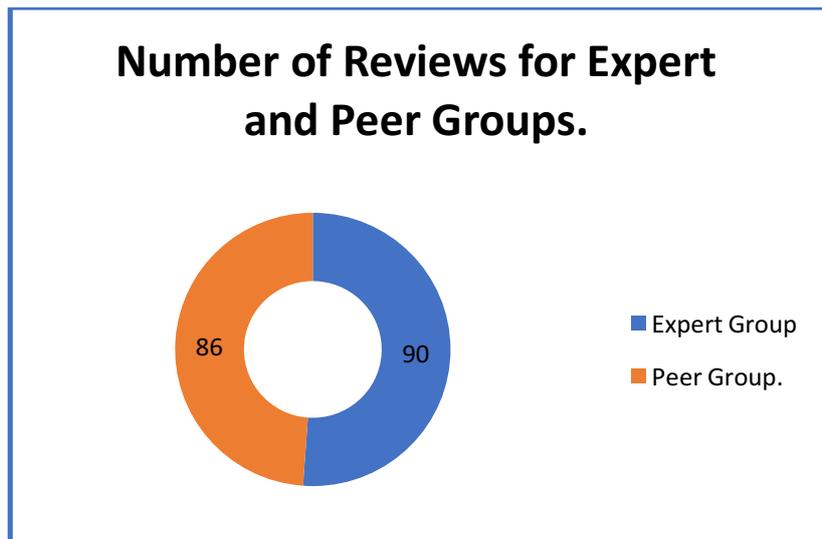


Figure 8. Numbers of reviews undertaken by expert and peer study participants

### **4.3 Distribution of Studies Reviewed.**

Not all ultrasound scans received equal numbers of reviews; i.e. not all participants completed all 22 study reviews. The number of reviews per participant is demonstrated graphically in 9. The mean number of ultrasound examinations reviewed was 9 (SD=4.1).

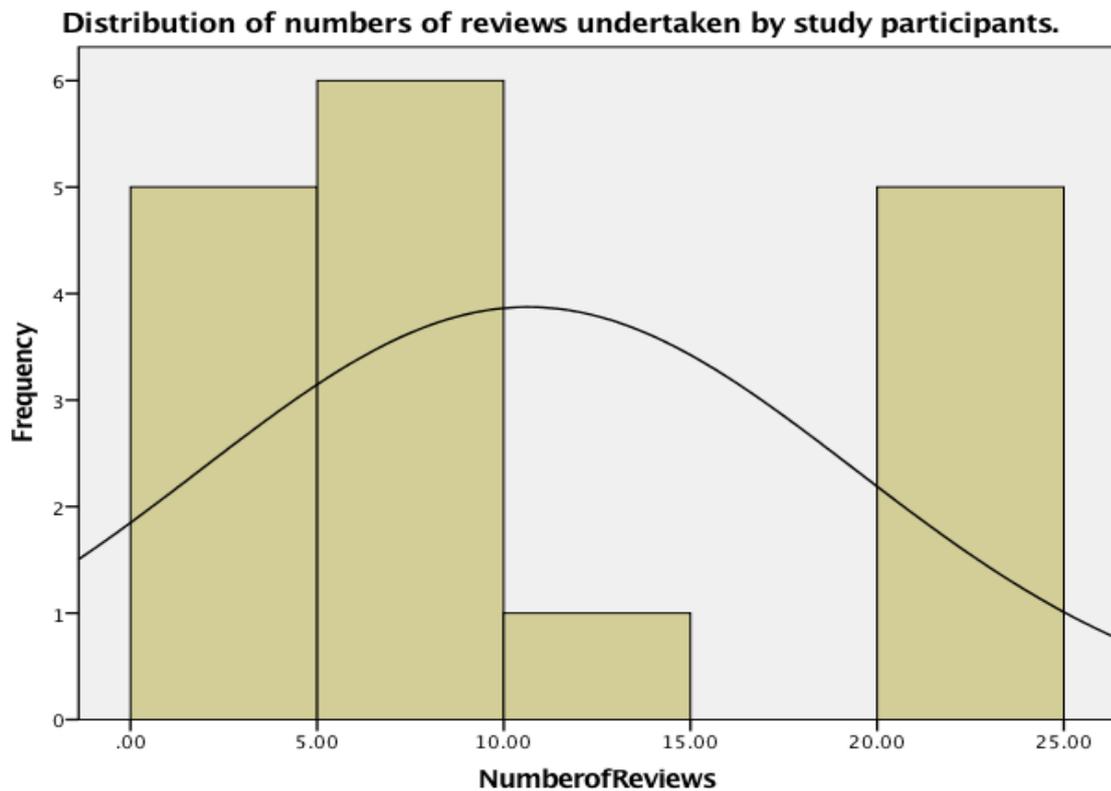


Figure 9. Distribution of numbers of Reviews completed by study participants.

### **4.4 Descriptive Statistics for the Quality Assessment Tools under Investigation.**

176 reviews of 22 ultrasound studies were performed by a total of 17 study participants. Not all participants reviewed all ultrasound studies (range of numbers

of examinations reviewed=1-22). Five study participants completed reviews of all 22 ultrasound examinations.

Descriptive statistics which summarise the overall scores for each of the quality measurement tools are given in **Error! Reference source not found.** 10.

| Table 10   |                   |                     |                           |
|--|-------------------|---------------------|---------------------------|
| <i>Summary of Scores for each Quality Rating Tool for 22 Ultrasound Scans.</i> |                   |                     |                           |
| <b>Quality Rating Tool</b>   | <b>Mean Score</b> | <b>Median Score</b> | <b>Standard Deviation</b> |
| Overall Study Score (0-10)   | 5.9               | 5.7                 | 2.2                       |
| Local Report Rating (1-5)  | 3.7               | 4.0                 | 1.2                       |
| AQP report rating tool (0-5)   | 3.4               | 3.6                 | 1.4                       |
| AQP image quality rating tool (1-5).   | 3.7               | 3.9                 | 1.1                       |

Table 10. Summary of Scores for each Quality Rating Tool for 22 Ultrasound Scans.

#### **4.5 Assessment of Distribution Curves for Normality.**

The data from the ultrasound rating tools (Table 10) was tested for normal distribution. This was to ascertain whether parametric statistical analysis was appropriate for this data. Normal distribution of study scores was tested following the guidance of Kim et al[62]. This suggests that for sample sizes between 50 and 300, normal sample distribution can be assumed with a z value (skewness/standard

error) of less than 3.29 for skewness. The degree of skewing and z score is given in **Error! Reference source not found.** for each of the quality rating tools under investigation.

Table 11

*Assessment of Normal Distribution of Scores for each Quality Rating Tool.*

| <b>Quality Rating Tool</b>           | <b>Min Score</b> | <b>Max Score</b> | <b>Mean</b> | <b>Standard Deviation</b> | <b>Skew</b> | <b>Skew SD</b> | <b>z score</b> | <b>Interpretation</b>    |
|--------------------------------------|------------------|------------------|-------------|---------------------------|-------------|----------------|----------------|--------------------------|
| Overall Study Score (0-10)           | 1                | 10               | 5.9         | 2.2                       | -.077       | .18            | -0.4           | Normal distribution      |
| Local Report Rating (1-5)            | 1                | 5                | 3.7         | 1.2                       | -0.71       | .18            | -3.9           | Non-normal distribution. |
| AQP report rating tool (0-5)         | 0                | 5                | 3.4         | 1.4                       | -0.89       | .18            | -4.9           | Non-normal distribution. |
| AQP image quality rating tool (1-5). | 1                | 5                | 3.7         | 1.1                       | -0.86       | .18            | -4.7           | Non-normal distribution. |

Table 11. Assessment of Normal Distribution of Scores For each Quality Rating Tool

Calculation of the z value for most of the reporting tools demonstrated significant negative skewness with z scores above the threshold of 3.29, except for the quality rating tool ‘Overall Score’ which was normally distributed. Logarithmic transformation of all data except for data from the rating tool ‘Overall Score’ was therefore undertaken. This resulted in resolution of the skewness and subsequent normal distribution of logarithmic data. Parametric testing was therefore appropriate for these transformed data. Table 12 shows the level of skewness after logarithmic transformation of the data.

| Table 12  |                  |                  |             |                           |             |                                |                |                                |
|---|------------------|------------------|-------------|---------------------------|-------------|--------------------------------|----------------|--------------------------------|
| <i>Assessment for Normal Distribution for Quality Rating Tools after log Transformation</i> |                  |                  |             |                           |             |                                |                |                                |
| <b>Quality Rating Tool</b>  | <b>Min Score</b> | <b>Max Score</b> | <b>Mean</b> | <b>Standard Deviation</b> | <b>Skew</b> | <b>Skew Standard Deviation</b> | <b>z score</b> | <b>Interpretation</b>          |
| Local Report Rating (1-5)   | .00              | .70              | .2943       | .23749                    | .056        | .183                           | 0.3            | <b>Normal log distribution</b> |
| AQP report rating tool (0-5)  | .00              | .78              | .3544       | .24071                    | -.101       | .183                           | -0.55          | <b>Normal log distribution</b> |
| AQP image quality rating tool (1-5).  | .00              | .70              | .3041       | .21022                    | -.085       | .183                           | -0.46          | <b>Normal log distribution</b> |

Table 12. Assessment of Normal Distribution of Scores for each Quality Rating Tool after log Transformation

#### **4.6 Assessment of Agreement Between Participants.**

Consistency and agreement of reviews between study participants was analysed by intra-class correlation coefficient (ICC). The interpretation of ICC has been described by Altman[63] and is given in

Table 13.

| Table 13  |                               |
|---|-------------------------------|
| <i>Values for Interpretation of Intra-class Correlation</i> |                               |
| <b>Value of ICC.</b>  | <b>Strength of Agreement.</b> |
| <0.2  | <b>Poor</b>                   |
| 0.21-0.40   | <b>Fair</b>                   |
| 0.41-0.60   | <b>Moderate</b>               |
| 0.61-0.80   | <b>Good</b>                   |

|           |                  |
|-----------|------------------|
| 0.80-1.00 | <b>Very Good</b> |
|-----------|------------------|

**Table 13. Strength of Agreement in ICC**

Because every study participant did not review all image sets, meaningful calculation of intra-class correlation was difficult. Five study participants completed all reviews of the available ultrasound imaging and agreement between these five reviewers has been calculated.

All 5 participants completing these reviews were sonographers/radiographers by profession. 3 of these sonographers were assigned to the 'expert-group' in this study and the remaining 2 were assigned to the 'peer-group'. Agreement between the 5 reviewers for each audit tool are given in table 14.

| Table 14<br><i>Agreement between Participants using different quality rating tools to assess 22 ultrasound examinations.</i> |  |                                |                       |
|--|--|--------------------------------|-----------------------|
| <b>Ultrasound Rating Tool</b>  | <b>Intra-class Correlation Coefficient</b> | <b>95% Confidence Interval</b> | <b>Interpretation</b> |
| Local Report Quality   | 0.2  | 0.05 - 0.43                    | Fair                  |
| Overall Score  | 0.3  | 0.13 - 0.52                    | Fair                  |
| AQP Report Quality   | 0.3  | 0.12 - 0.53                    | Fair                  |
| AQP Image Quality  | 0.38                                       | 0.19 - 0.61                    | Fair                  |

**Table 14. Agreement between 5 participants using different quality rating tools to assess 22 ultrasound examinations.**

Unfortunately, due to the small number of fully completed data sets, it is not possible to examine agreement between study participants in more detail (for example, to see whether agreement is better between experts or non-experts) as confidence interval would be too wide to give meaningful results. However, the data does demonstrate only 'fair' agreement between study participants in scoring of image and report quality of 22 selected ultrasound examinations. There is no large difference in levels of agreement between any of the rating tools used.

#### **4.7 Effect of Participant Characteristics on Quality Scores Given.**

To determine whether there were systematic differences in quality scores with respect to individual participant characteristics, it was necessary to regress the data. This was an attempt to extract whether each of the participant characteristics independently influenced the scores given. If this were the case it would be necessary to estimate the magnitude of this effect.

After logarithmic transformation, each of the quality scores given demonstrated a normal distribution and therefore regression was an appropriate method of analysis. How each of the participant characteristics (clinical background, clinical grade, years of clinical experience, age of equipment within a reviewer's department, peer/expert participant) contributed to the scores for each rating tool could therefore be measured. Dummy variables were utilised to facilitate analysis of categorical data (clinical grade) within the regression.

The null hypothesis was that none of these independent variables would have any predictive effect on the scores given by study participants. The null hypothesis was considered disproved at a p value of 0.05 or less.

It was acknowledged that use of log transformed data would limit analysis of the magnitude of effect of the different variables, when significant systematic differences between groups were found. Assessment of the  $\log\beta$  coefficient against the possible range of scores was used to give an estimated effect size for logarithmically transformed data.

#### **4.8 Effect of Individual Factors on Ultrasound Examination Ratings.**

##### 4.8.1 Quality Rating Tool: Overall Score.

Table 15 demonstrates the effect of participant characteristics on the scores given for the rating tool 'Overall Score'. Sonographer AfC banding has a statistically significant effect ( $p=.021$ ) with band 8a sonographers likely to give a higher score than either band 7 or band 8b sonographers. The magnitude of this effect appears to be clinically significant with ultrasound examinations rated by band 8a sonographers scoring an estimated 1.6 points higher (out of 10) than the same examinations rated by sonographers of other clinical grades. There are no systematic differences for other participant variables.

| <b>Table 15</b>   |                                    |                |                     |                                      |       |
|---|------------------------------------|----------------|---------------------|--------------------------------------|-------|
| <b><i>Regression for Quality Rating Tool: Overall Score</i></b> |                                    |                |                     |                                      |       |
| <b>Participant Characteristic</b>                               | <b>Unstandardised Coefficients</b> |                | <b>Significance</b> | <b>95% Confidence interval for B</b> |       |
|   | B                                  | Standard Error | Significance        | Lower                                | Upper |
| Band 7  | 0.531                              | .648           | .414                | -.748                                | 1.810 |

|                       |       |      |      |        |       |
|-----------------------|-------|------|------|--------|-------|
| Band 8a               | 1.618 | .695 | .021 | .246   | 2.991 |
| Band 8b               | -.064 | .663 | .923 | -1.373 | 1.245 |
| Expertise             | -.096 | .397 | .810 | -.879  | .688  |
| Equipment Age (years) | -.036 | .108 | .741 | -.250  | .178  |
| Experience (years)    | .005  | .022 | .807 | -.038  | .048  |

Table 15. Regression for Quality Rating Tool: Overall Score

#### 4.8.2 Local **Report** Quality Rating Tool

Table 16 demonstrates the effect of participant characteristics on the scores given for the ultrasound report rating tool, currently in local use. The analysis demonstrates that lower grade sonographers (band 7) are likely to rate clinical ultrasound reports more favourably when using this tool than sonographers of a higher grade ( $p=.025$ ). The magnitude of the increase in scores offered by band 7 sonographers was  $\log 0.16$ . Given that the log range of potential scores is 0 – 0.7, this is a potentially significant effect with the potential for band 7 sonographers to systematically increase the score of ultrasound scans into a higher (more favourable) category. However, the confidence intervals are wide (.02 - 2.9) and the actual effects of this are difficult to confirm with confidence.

| Table 16.  |                             |                |              |                               |       |
|--|-----------------------------|----------------|--------------|-------------------------------|-------|
| <i>Regression for Local Report Quality Rating Tool</i> |                             |                |              |                               |       |
| Participant Characteristic                             | Unstandardised Coefficients |                | Significance | 95% Confidence interval for B |       |
|  | B                           | Standard Error | Significance | Lower                         | Upper |
| Band 7   | .156                        | .069           | .025         | .020                          | .292  |

|                       |       |      |      |       |      |
|-----------------------|-------|------|------|-------|------|
| Band 8a               | .089  | .071 | .208 | -.050 | .229 |
| Band 8b               | .084  | .074 | .261 | -.063 | .230 |
| Expertise             | .079  | .042 | .063 | -.004 | .163 |
| Equipment Age (years) | -.022 | .012 | .059 | -.045 | .001 |
| Experience (years)    | .003  | .002 | .164 | -.001 | .008 |

Table 16. Regression for Local Report Quality Rating Tool

#### 4.8.3 AQP Report Quality Rating Tool.

None of the participant characteristics under investigation had any significant effect on the scores given for ultrasound scan reports when rated using the Report Quality Rating Tool required under the AQP system. Regression figures are given in Table 16.5.

| Table 16.5  |                             |                |              |                               |       |
|---|-----------------------------|----------------|--------------|-------------------------------|-------|
| <i>Regression for the AQP Report Quality Rating Tool.</i> |                             |                |              |                               |       |
| Participant Characteristic                                | Unstandardised Coefficients |                | Significance | 95% Confidence interval for B |       |
|   | B                           | Standard Error | Significance | Lower                         | Upper |
| Band 7  | 0.73                        | 4.7            | 0.88         | -8.6                          | 10.1  |
| Band 8a   | 1.1                         | 3.2            | 0.74         | -5.2                          | 7.4   |
| Band 8b   | 0.2                         | 1.7            | 0.9          | -3.1                          | 3.6   |
| Expertise   | -0.4                        | .259           | .088         | -0.9                          | 0.7   |
| Equipment Age (years)                                     | 0.01                        | .07            | 0.90         | -0.1                          | 0.1   |
| Experience (years)  | 0.01                        | 0.01           | 0.47         | -0.02                         | 0.04  |

Table 16.5. Regression for the AQP Report Quality Tool.

#### 4.8.4 AQP Image Quality Rating Tool.

This quality rating tool looked specifically at the quality of the ultrasound images produced and stored, rather than the written report generated by such images.

Table 17 shows that there are two participant characteristics which have a statistically significant effect on the scores given when using this rating tool. Expert participants give higher scores for image quality than the peer group participants within this study ( $p=0.03$ ). The length of clinical experience of the study participant also has an effect with higher scores being given by those of longer experience ( $p=0.03$ ).

While statistically significant, the *clinical* significance of the effect of expert compared with peer group participant is doubtful. The effect of utilising an expert (rather than peer) reviewer raises the score by  $\log 0.08$  (0.008 – 0.153). Given that the log range of potential scores for image quality ranges from 0 – 0.7, it is unlikely that the scores offered by expert reviewers would have any significant clinical effect on an audit program when compared with the scores given by a peer group.

The effect of length of clinical experience of participants is statistically significant and the clinical significance is also potentially significant. The model demonstrates that for each year of clinical experience, the image quality score increases by  $\log 0.004$ . In theory, given that the range of log scores is 0 - 0.7, 40 years of clinical experience potentially raises the image quality score by 23% when compared with a rater of no clinical experience.

In practice, it is unlikely that the effect of clinical experience is linear in nature. It is more probable that the effect of clinical experience is more pronounced early on in a rater’s career, plateauing at a level when clinical maturity is reached. Unfortunately, this study does not have the statistical power to model such non-linear effects.

While this study confirms that those of greater clinical experience award statistically higher scores for image quality when using the ‘AQP Image Quality’ rating tool, the precise effect of this interaction is difficult to model due to the limitations of this data.

| Table 17   |                             |                |              |                               |       |
|--|-----------------------------|----------------|--------------|-------------------------------|-------|
| <i>Regression for the Image Quality Rating Tool.</i> |                             |                |              |                               |       |
| Participant Characteristic                           | Unstandardised Coefficients |                | Significance | 95% Confidence interval for B |       |
|  | B                           | Standard Error | Significance | Lower                         | Upper |
| Band 7   | .07                         | 0.06           | 0.23         | -0.05                         | 0.19  |
| Band 8a  | -0.01                       | .065           | 0.89         | -0.14                         | 0.12  |
| Band 8b  | 0.08                        | .062           | 0.2          | -0.04                         | 0.20  |
| Expertise  | 0.08                        | .037           | .031         | 0.01                          | 0.15  |
| Equipment Age (years)                                | 0.007                       | .010           | 0.51         | -0.01                         | 0.03  |
| Experience (years)                                   | 0.004                       | .002           | .03          | .00                           | .008  |

Table 17. Regression for the Image Quality Rating Tool.

#### **4.9 Correlation between Quality Measurement Tools.**

All ultrasound rating tools under investigation were used on the same group of ultrasound cases. i.e. study participants used a range of ultrasound rating tools to rate the *same* group of ultrasound examinations. A high degree of correlation would

therefore be expected between quality assessment tools (as each was used on the same group of ultrasound imaging). However, it was felt important to confirm correlation between the different ultrasound rating tools to provide evidence that the rating tools were measuring the same thing (ultrasound examination quality). This was assessed using a 2-tailed correlation coefficient with Spearman’s test to assess for statistical significance (table 18).

The null hypothesis was that there would be no correlation between the different ultrasound rating tools. This was considered disproved at a p value of 0.01 or below.

| Table 18.   |                     |                                    |                      |                           |                          |
|---|---------------------|------------------------------------|----------------------|---------------------------|--------------------------|
| <i>Correlation between Ultrasound Quality Rating Tools.</i> |                     |                                    |                      |                           |                          |
|   |                     | <b>Report Quality (local tool)</b> | <b>Overall Score</b> | <b>AQP Report Quality</b> | <b>AQP Image Quality</b> |
| <b>Report Quality (local tool)</b>                          | Pearson Correlation | 1                                  | 0.51                 | 0.76                      | 0.38                     |
|   | Significance        |                                    | <i>&gt;0.01</i>      | <i>&gt;0.01</i>           | <i>&gt;0.01</i>          |
| <b>Overall Score</b>  | Pearson Correlation | 0.510                              | 1                    | 0.610                     | 0.591                    |
|   | Significance        | <i>&gt;0.01</i>                    |                      | <i>&gt;0.01</i>           | <i>&gt;0.01</i>          |
| <b>AQP Report Quality</b>                                   | Pearson Correlation | 0.756                              | 0.610                | 1                         | 0.417                    |
|   | Significance        | <i>&gt;0.01</i>                    | <i>&gt;0.01</i>      |                           | <i>&gt;0.01</i>          |
| <b>AQP Image Quality</b>                                    | Pearson Correlation | 0.378                              | 0.591                | 0.417                     | 1                        |
|   | Significance        | <i>&gt;0.01</i>                    | <i>&gt;0.01</i>      | <i>&gt;0.01</i>           |                          |

Table 18. Correlation Between Ultrasound Quality Rating Tools.

Each of the rating tools assessing ultrasound report quality are correlated as expected.

The rating tools for AQP *image* quality and rating tools examining ultrasound *report* quality would not necessarily be correlated as it would be entirely possible to have good image quality and poor report ultrasound quality and vice-versa. This study does show that image quality and report quality scores are related. Study participants tend to score image quality and report quality in a similar fashion for an individual ultrasound examination.

## **Chapter 5**

### **Discussion of Quantitative Results.**

## **Chapter 5.**

### **Discussion of Quantitative Results.**

**This chapter seeks to both summarise and interpret the quantitative results generated by this project. It aims to embed the results firmly within the context of knowledge from existing literature.**

#### **5.1. Introduction.**

The results of this project have uncovered some important issues which need to be addressed in terms of application to clinical practice:

##### **5.1.1 Appropriateness of current quality assurance methods.**

Is the current methodology of quality assurance in diagnostic ultrasound appropriate? This study has highlighted that even when participants were given the same ultrasound images and reports and used the same audit tools to rate these ultrasound examinations, inter-rater agreement was only 'fair'. Given this level of agreement, is current methodology suitably robust to fulfil requirements?

##### **5.1.2 Systematic effects of participant characteristics.**

There is little evidence to suggest that systematic effects of the various participant characteristics are easy to understand or consistent in terms of eventual quality scores given. These characteristics included clinical expertise, length of clinical experience, clinical grade and age of equipment used. Because systematic differences cannot be consistently demonstrated, effective modelling of the effects of these different factors is complex.

#### 5.1.3 Free-text comments from study participants.

The website possessed the functionality to enable study participants to leave free-text comments on the ultrasound imaging that they had rated. This was optional, but most participants chose to leave comments, some of which were extensive. This was an unexpected, but rich source of qualitative data. Although this study was predominantly quantitative in nature, the quantity and depth of these comments also warranted further evaluation. The methodology for this is described in Chapter 6.

#### 5.1.4 Use of internet for feedback.

The internet provides a useful forum for offering feedback to ultrasound practitioners who are geographically separate. With simple IT infra-structure, the logistics for this are relatively straightforward. It is important to determine whether giving remote (and possibly) anonymous feedback result in quantifiable, positive outcomes. The potential effects of internet use in this manner requires further evaluation. This will be explored further in Chapter 7.

#### 5.1.5 Multidisciplinary nature of research.

This study was intended as multidisciplinary in nature with participants drawn from a range of background professions. However, most responses came from sonographers, rather than other professional groups. This opens questions as to how quality assurance processes should be carried out and which professions would be best placed to undertake this work.

## 5.2 Inter-rater Reliability of Ultrasound Reviews.

This study demonstrates sub-optimal levels of agreement between reviewers when undertaking reviews of diagnostic ultrasound examinations including examinations of the abdomen, urinary tract and female pelvis. Given that retrospective review of static imaging is currently the most common method of ultrasound quality assurance, these results are a little disturbing. Ultrasound examinations are being rated using methodology and quality assurance tools where inter-rater agreement is only classed as 'fair' when tested objectively.

### 5.2.1 Use of ultrasound image scoring tools.

Reviewers were asked to review and rate the *same* imaging ultrasound studies, displayed on a standard personal computer screen with the *same* level of detail regarding clinical presentation and patient demographic data. The use of the internet in displaying images ensured standardisation as much as reasonably possible in terms of reviewing conditions. However, even after reviewing the same imaging, written ultrasound reports and patient histories, when independently scored by different ultrasound practitioners the inter-rater agreement was only 'fair'. It is therefore important to consider whether this was due to some defect or flaw within the imaging scoring tools or whether such ambiguity was inherent in the method of quality assurance itself.

It could be argued that the design and wording of the audit tools were too ambiguous to yield high levels of inter-rater agreement. Better development of the rating tools, use of more detailed tools, or providing training for study participants in the use of these tools may therefore have yielded more consistent results between individual

participants. However, the tools were to be used for rating a wide range of ultrasound imaging. They were deliberately designed to be reasonably ambiguous to facilitate application against a wide range of ultrasound imaging and by a wide range of individuals of varying experience, clinical grade and professional background. Use of very detailed rating tools would limit the scope of use of these tools and therefore the scope of a quality assurance program in ultrasound. The tools therefore needed to be sufficiently flexible to be adaptable to most clinical situations.

A previous study[39] demonstrated that experienced ultrasound practitioners tend to use heuristic methods of ultrasound image evaluation. Even when given very detailed instructions and image review tools, they tend to 'reverse-engineer' their responses to give a final score which they feel is most appropriate. This study set out to evaluate several different ultrasound image and report evaluation tools using a variety of scales. Some of these rating tools are already in clinical use as part of contractual obligations to provide quality assurance data on the current quality of ultrasound imaging and reporting[38]. Others have been developed by individual ultrasound units for local use[26].

Despite a plethora of different quality assessment tools and scoring methods, there is no real difference in the level of inter-rater agreement when these different tools are objectively tested. This suggests that the lower than optimum levels of participant agreement is embedded within the *method* of quality assurance itself,

rather than simply being due to variance or deficiency of the audit tools currently in use.

#### 5.2.2 Consequences of High Inter-Rater Variability.

The lower than expected level of participant agreement when assessing the quality of ultrasound examinations has not been previously reported. Diagnostic ultrasound is renowned for its subjectivity in performance and interpretation[43, 64], yet there is very little literature to evaluate how, or indeed whether the different quality assurance methods manage this subjectivity. This is surprising as important clinical and commercial decisions are made based on (assumed) accuracy and quality of ultrasound results. There seems to have been little meaningful attempt to evaluate and understand the methods by which such assumptions are arrived at.

Gaining an understanding of the reasons, effects and possible implications of these levels of participant agreement is difficult due to the absence of existing literature in this very specialist field. To gain a deeper understanding of the reasons and potential consequences of these findings, it is necessary to find an analogous situation which has been more comprehensively researched, and from which comparisons and conclusions may be drawn.

The peer review process for publication of scientific literature provides an analogous situation. It is therefore reasonable to explore whether the existing knowledge around the scientific peer review process can be extrapolated to give some context

to the implications of the results from this research project in the specialist field of diagnostic ultrasound.

### 5.2.3 Peer Review of Scientific Literature.

Peer review of the scientific process and reporting of results provides a cornerstone in assessment of the quality and validity of scientific literature. While the imperfections of such a process have long been known, it is generally acknowledged as the 'least imperfect way of upholding the quality of scientific publications'[65].

Peer review prior to scientific publication also exhibits high levels of inter-reviewer variation with Cohen Kappa scores of between 0 and 0.4[57] although some argue that the research and statistical methodology in the assessment of this is flawed[57]. However, there is a general consensus that peer review is at best, imprecise with some arguing that it is 'little better than a dice roll'[66].

The similarities between peer review of the scientific literature and retrospective review of the quality of diagnostic ultrasound studies is striking. Peer review of the literature and retrospective review of ultrasound imaging both place a high degree of subjectivity on the part of the reviewer. The tools used are (necessarily) imprecise to accommodate and be relevant to the large variation in subject matter presented. Unsurprisingly, such conditions do not lend themselves well to high degrees of precision and inter-rater agreement. Whether this is necessarily a bad thing, or simply a fact of life is a moot point.

Some commentators argue that the imprecise and variable nature of reviews in the scientific literature may be an asset. Differences between reviewers encourage opposing and differing opinions to be taken into account[67]. Homogenous views which do not seek to challenge the 'status quo' have been criticised for stifling innovation and leading to rather conservative attitudes in terms of introduction of new knowledge and technologies[57, 68]

#### 5.2.4 Acceptability of Inter-Rater Variation.

How much variation is acceptable between reviewers? Clearly, any process which seeks to inform and benchmark the quality of work offered (whether for diagnostic ultrasound quality assurance purposes or peer review of the scientific literature) needs to be sufficiently robust to maintain the credibility of the process itself. However, making a prescriptive judgement as to an acceptable level of inter-rater variation is exceedingly difficult. By the very nature of the review process, attempting to prescribe a specific cut-off figure for agreement in such a subjective area is unlikely to be successful.

In the peer review process for publication, the consensus seems to be that current levels of inter-rater agreement between scientific reviewers are too low and require improvement to improve the credibility of the process[57, 69-71]. Does a similar consensus apply to quality assurance procedures in diagnostic ultrasound? This is an important consideration and in trying to find an answer, it is important that the differences between quality assurance of diagnostic ultrasound examinations and reviewing of scientific writing for publication are also recognised. The most important

of these considerations is to determine the objectives and end-points of such a process.

#### 5.2.5 Objectives for peer review and quality assurance.

The dominant reason for peer review in the scientific literature is to assess whether a paper is of sufficient merit for publication within a given journal[72]. While the peer process can also help to improve and 'polish' a paper prior to publication, this is a secondary function.

In diagnostic medical ultrasound, there is likely to be more plurality in the objectives of a quality assurance program. It *may* simply function as an assurance of quality during the commissioning process. However, of equal or greater importance is the identification of areas of weakness which can be used to inform and drive quality improvements[26]. Thus, a quality assurance program may be either an evolving process of quality improvement or a quality-control process focused on an end-product, or a combination of both. Individual ultrasound departments need to be very clear which of these functions their quality assurance program is to serve when trying to decide what level of inter-rater agreement is acceptable.

There is anecdotal evidence that at least some ultrasound providers are using these quality assurance processes solely as a mechanism of quality control rather than quality improvement[73], i.e. that this method of quality assurance is being undertaken simply to demonstrate compliance with existing or potential commissioned contracts. In this case, the degree of inter-rater agreement found in this study would seem to be unacceptably low. That some ultrasound providers are

using a single practitioner (usually a consultant radiologist) to provide evidence of satisfactory standards raises the possibility of unacceptably high variation in results. The results obtained from such practices are liable to demonstrate as much about the individual reviewing practitioner as about the ultrasound imaging being appraised. Any quality control data obtained under these conditions should be treated with caution.

Where diagnostic ultrasound studies are being appraised for quality improvement (rather than quality control) purposes, a higher degree of inter-rater variation is acceptable and may even be beneficial. Even where there are high levels of disagreement between ultrasound practitioners, this can begin to stimulate debate as to what constitutes a good or bad ultrasound examination and report. Such discussion is likely to be very helpful in terms of defining, agreeing and crystallising standards of working practice among a cohort of sonographers. The creation of agreed common principles both in undertaking ultrasound examinations and scoring of these ultrasound examinations for quality assurance purposes is likely to have beneficial outcomes in terms of department clinical standards.

#### 5.2.6 Can Inter-Rater Agreement be Improved?

It has been suggested that inter-rater agreement could be improved by reducing the variation inherent during an ultrasound examination. By standardising the types of images recorded for a given ultrasound examination and by standardising the reporting of such ultrasound examinations (e.g. by use of reporting templates), quality review tools could be designed and applied more objectively, thereby reducing the level of variation between reviewers. Such a practice has been

explicitly warned against by the Royal College of Radiologists and College of Radiographers[26]. There is concern that such an approach has the potential to undermine diagnostic ultrasound's greatest strength. i.e. the ability to examine normal and abnormal anatomical structures from a number of positions and angles to best demonstrate normal and pathological appearances, thereby contributing to accurate and timely patient diagnosis. A very 'protocol-driven' approach may help to reduce variation in assessment of examination quality, but the cost of such an approach is to inhibit the fullest potential of diagnostic ultrasound to be realised.

Evaluation of the literature describing scientific peer review show that some attempts have been made to achieve more standardisation in determining the merits and deficits of scientific papers submitted for publication. The importance of having more than one reviewer examining any work is stressed to minimise the effects of natural variation among individual reviewers. All journals utilise more than one reviewer for each scientific paper to accommodate this.

Anecdotal evidence suggests that the use of multiple reviewers in quality assurance of medical ultrasound is not current practice for many diagnostic ultrasound departments. Although current, detailed knowledge of the different mechanisms of quality assurance in current use are not yet available, anecdotal evidence suggests that many centres undertaking this work utilise a single reviewer for quality assurance purposes. This has the potential to give flawed data on the measurement of quality of the ultrasound imaging reviewed.

### 5.2.7 Can Training in Reviewing Help?

Can provision of training help to improve the consistency of reviews of ultrasound imaging and reporting? There is little in the scientific literature in this specialist area to indicate what difference training would make. This study demonstrates that even among experts, there is little agreement about what constitutes a 'good' ultrasound examination and report. By extension, there is likely to be variation in the content and provision of training programs designed to address this.

It is helpful to use the scientific peer review process to understand the effects of a training program on the quality of subsequent reviews, and the likely resources required to effect such a change. There has been some primary research which has attempted to evaluate and reduce the level of variation between reviewers during the peer review process for publication by introduction of a training program.

Smith [74] described a study that attempted to improve quality and reduce the variation inherent in peer review of scientific papers for publication. This described the provision of computer-based and face-to-face training in the reviewing process to an intervention group compared with a control group receiving no training when considering papers for publication. Although there was a small improvement after training, the authors concluded that the improvement was not sufficiently large to merit provision of a training program. This has been confirmed by other studies[69, 75]which showed no significant or long-term specific effect of training on the subsequent quality of reviews given.

Other strategies for reducing variation between reviewers such as providing mentoring for new reviewers also appear to have little effect on the subsequent quality of reviews given[76]. Implementing more stringent selection processes for reviewers based on clinical grade and length of experience[77] also do not improve the level of inter-rater agreement, particularly for those papers which are at variance with an experienced reviewer's epistemological standpoint[78]. Training, mentoring and experience are therefore of limited help, but it remains important to try and find other measures which can be put in place to attempt to improve the quality of the review or quality assurance process.

#### 5.2.8 Blinding of reviews.

The most common current methodology in peer-review for scientific publication is that of double blinding. i.e. neither reviewers nor authors are aware of the identity of the other participants within the process. This is analogous to this research project, in which participants were required to give anonymous reviews on ultrasound studies without knowledge of where and by whom these studies were performed. It is possible that the very nature of blinding (and asking reviewers to undertake reviews of ultrasound studies performed outside of their institution) gives rise to reviews which are unnecessarily harsh, because it is easy to forget the pressures that sonographers are working under, and the difficulties that some patients can present to obtaining high quality imaging examinations. It is important to establish whether the degree of reviewer blinding affects the quality or content of reviews given.

There is some literature that studied the effect of unblinding in the scientific peer review system. This system of unblinding may be limited in that reviewers are made

aware of the identity of the author, or a completely open system such that authors and reviewers are all aware of each other's identities. Unblinding of reviewers does not seem to have any effect on the quality of the review given[79-81], but there is evidence that reviewers are sometimes uncomfortable in having their anonymity revealed which may result in a reduction the number of reviewers willing to undertake such reviews[82].

#### 5.2.9 Scientific Peer Review knowledge and its application to quality assurance of ultrasound imaging.

In summary, there is currently no primary research which looks at inter-rater agreement in quality assurance of diagnostic ultrasound imaging. This is the first study to specifically address this area to the best of the principal researcher's knowledge. The novelty of this study and consequent absence of existing literature to provide context does raise some challenges in terms of evaluation of the significance and implications of the results obtained.

There are similarities between the peer review process for scientific publication and review of ultrasound imaging for quality assurance purposes. Both processes require non-specific and broad assessment tools to capture reviews from a wide range of sources and subjects. The subjects themselves (scientific papers or diagnostic ultrasound studies) are highly variable in nature and the assessment of these is therefore highly subjective. Under these conditions, it is perhaps unsurprising that the levels of inter-rater agreement are only 'fair' for both subject matters.

There are no easy or quick methods of improving the levels of inter-rater agreement. Training, mentoring and unblinding of reviewers appear to have little effect and may be counterproductive, particularly in terms of finding external reviewers willing to undertake unblinded assessments of diagnostic ultrasound imaging.

The levels of inter-rater agreement do give rise to some concern, particularly in providing robust evidence of adequate standards during commissioning procedures. Therefore, in terms of using this method of quality assurance within an individual ultrasound unit, a judgement needs to be taken as to whether the levels of inter-rater agreement found within this study still make a quality assurance program worthwhile.

If the aim is to demonstrate compliance to a minimum level of service quality to fulfil agreed service contracts, then this method of quality assurance is likely to be inadequate. The data has suggested that the individual reviewer has at least as much impact as the actual ultrasound imaging under review in terms of the scores assigned to an ultrasound examination, regardless of the rating tools used. Thus, demonstrating high levels of compliance and competence, at least theoretically depends on selection of the 'right' reviewer.

This is a system which is open to abuse. There is evidence in the literature that pre-selection of reviewers for scientific papers gives rise to excessively positive reviews, even when unmerited[83]. It would be entirely possible for unscrupulous ultrasound providers to undertake ultrasound quality assurance, merely as a 'box-ticking exercise' to provide evidence of compliance with contractual obligations. This could theoretically be achieved by selection of the 'right' reviewer without any meaningful intention to maintain and improve service quality. The method of appointing a single

reviewer for the purposes of quality control for commercial practices is therefore to be discouraged.

A more meaningful objective is to use quality assurance processes to drive continual improvements in the quality of ultrasound services offered. It could be argued that the stated levels of inter-observer agreement are of less importance if a quality assurance program is implemented in this context. There is evidence in the literature that while open-panel discussion does not improve the inter-rater variation between scientific reviews [84], this approach does at least allow a consensual approach to be taken in reviewing of ultrasound imaging.

Stimulating debate among ultrasound practitioners may be very helpful in identification and management of areas of sub-optimal clinical practice. In this context, high levels of inter-rater agreement may actually be counterproductive in that it may stifle genuine, constructive debate. Dissenting voices with a differing, yet equally valid viewpoint may not be heard resulting in continued reliance on the status quo, rather than a genuine commitment to service improvement. The importance of recognition and challenging existing ways of working, especially in areas where improvements are desirable, should not be underestimated.

From this viewpoint, several reviewers who have differing opinions may be helpful in debating and coming to a consensus about what constitutes poor, acceptable and good practice with subsequent improvements where necessary. Clearly, where several reviewers are used in the assessment of ultrasound clinical quality, the individual characteristics of those reviewers could be an important consideration

when assessing for potential systematic effects between reviewers on the eventual quality scores given.

### **5.3 Systematic Effects of Participant Characteristics on Quality Scores given.**

This study sought to determine whether individual participant characteristics had a systematic effect on scores given when reviewing the ultrasound imaging and reports chosen for this study. A summary of the effects of the participant characteristics under investigation are given in Table 19.

| Table 19.<br><i>Summary of the Effect of Participant Characteristics on the Scores Given.</i> |  |   |   |
|---|--|---|---|
|   | <b>Image Quality</b>   | <b>Report Quality</b>   | <b>Overall Score.</b>   |
| <b>Participant Clinical Grade</b>   | Clinical grade has no effect on image quality scoring.   | Band 7 sonographers rated studies more highly using the local report rating tool. | Band 8a sonographers gave a higher score (1.6) than band 7 or band 8b sonographers. |
| <b>Participant Experience</b>   | More experienced staff are likely to rate image quality more highly. Requires further study to model the precise effect. | No effect demonstrated  | No effect demonstrated.   |
| <b>Equipment Age</b>  | No effect demonstrated.  | No effect demonstrated  | No effect demonstrated.   |
| <b>Expertise.</b>   | Expert group give statistically higher scores than the peer group. Probably not clinically significant                   | No effect demonstrated  | No effect demonstrated.   |

Table 19. Summary of effect of participant characteristics on scores given

### 5.3.1 Effect of Professional Grade.

Clinical grade has some effect on quality scores given, although there is no clear trend for increasing or decreasing clinical grade. Band 8a practitioners tend to give higher overall scores than band 7 or band 8b sonographers when using a continuous analogue scale which could be clinically significant with an overall increase in score of 1.6 (out of 10) when compared with other clinical grades of staff.

This result was not replicated when assessing the AQP or local quality assessment tools for measurement of report quality (rather than overall score). Band 7 sonographers gave higher scores for report quality using the local report rating tool than those of a higher grade. However, this difference was not repeated when utilising the AQP report rating tool.

The report quality rating tool used therefore has some effect on how reviewers of different clinical grades report quality in diagnostic ultrasound. To try and determine the reasons for this requires a review of both tools to define their differences and similarities. The tools in question are given in Tables 20 and 21.

| Table 20                                   |   |
|--|---|
| <i>Assessment Tool: AQP Report Quality</i> |   |
| <b>Score</b>                               | <b>Description</b>  |
| <b>5</b>                                   | Content and structure optimal   |
| <b>4</b>                                   | Essence of report satisfactory – slight modification of emphasis or advice  |
| <b>3</b>                                   | Report satisfactory but additional differential diagnosis or advice could have been provided. Unlikely to lead to patient harm  |
| <b>2</b>                                   | Discrepancy of measurement or interpretation. No immediate harm to patient but requires amended report  |
| <b>1</b>                                   | Unnecessary advice leading to inappropriate further investigation. For example: “can’t exclude malignancy” in clearly defined condition leading to invasive test or one involving ionising radiation when unnecessary. Inappropriate follow up recommended leading to downstream costs and patient anxiety. |
| <b>0</b>                                   | Poor report with risk of inappropriate management pathway   |

Table 20. Assessment Tool; AQP report quality.

| Table 21                          |   |
|-----------------------------------|---|
| <i>Local Report Quality Tool.</i> |   |
| <b>Score</b>                      | <b>Description</b>  |
| <b>5</b>                          | Complete agreement or only very minor change in wording or focus.           |
| <b>4</b>                          | Report accurate and of high quality but additional comment required.        |
| <b>3</b>                          | Report accurate but additional differential diagnoses have not been offered |
| <b>2</b>                          | Disagreement with image interpretation                                      |
| <b>1</b>                          | Clinical Question not answered or cannot be inferred from report.           |

Table 21. Local report quality tool.

Although the assessment tools were designed to measure the same thing (quality of the ultrasound report), there are some fundamental differences between the tools.

- The local report quality tool is on a 5-point scale whereas the AQP quality tool is on a 6-point scale. This may influence participant responses with more options to choose from when using the AQP report quality tool than the local report quality tool.
- The wording and intended meaning of both rating tools is similar between assigned scores of 2 and 5. However, there is a major difference between the Local and AQP report rating tools for ultrasound examinations which are assigned lower scores; i.e. for those ultrasound reports which are rated poorly by study participants. A score of 1 (Local rating tool) is given when the clinical question has not been answered. In contrast, when using the AQP tool, a score of 1 is given where the clinical advice offered is inappropriate.
- The AQP tool also offers an additional score of 0 where the report is deemed to be 'poor with risk of inappropriate patient management'. This terminology is ambiguous as the term 'poor' could be interpreted in different ways by different reviewers. The limited number of participants within this study do not permit a detailed analysis of whether reviewer grade is of more influence when assessing ultrasound reports of poorer, rather than higher quality. It is possible that the differences between study participants are not evenly

distributed across the quality spectrum of scores given. Differences between participants may be more marked at the lower end, rather than the higher end of the scales. Unfortunately, the data does not support this level of detailed analysis and is an area in which further research may be helpful.

### 5.3.2 Effect of Length of Participant Experience.

Length of participant experience has no effect on the scores given for ultrasound report quality. There was however, a detectable, systematic effect on the scoring of *image quality* according to the length of participant experience. More experienced participants tended to offer higher scores for image quality when compared to those of less experience. The effect of participant experience per year is small, but over the course of a 40-year career, the effect may be sizeable. This does however assume a linear relationship between length of experience and size of effect which, in practice is unlikely. It is more likely that the effect is greater early on in a participant's career plateauing at a point before the end of a participant's career with little subsequent increase in effect size.

The effect of length of participant experience therefore requires further data for accurate modelling. The reason for the differences between participants of different length of experience on image quality scores is thought most likely due to a more pragmatic approach in assessing image quality taken by more experienced reviewers. The quality of the ultrasound image is heavily influenced by such factors as patient obesity and amounts of internal bowel gas, as well as by the skill and experience of an individual ultrasound practitioner. Thus, while image quality is partially dictated by the intrinsic ability of the ultrasound practitioner, there are some

patients in whom high quality imaging is simply not possible, regardless of the skill of those undertaking the ultrasound examination. It is likely that this is more readily recognised by more experienced ultrasound practitioners who take a pragmatic approach to the 'difficult to image' patient when compared with their less experienced counterparts. More experienced practitioners could be more 'forgiving' of sub-optimal imaging because of their greater appreciation of the inherent difficulties of diagnostic ultrasound in some patients.

### 5.3.3 Effect of Equipment Age.

This characteristic was factored into the study to determine whether those participants accustomed to using more up-to-date equipment would rate studies differently when compared with those using older equipment. The rationale behind this was to determine whether the rating of an ultrasound study is affected by what the study participants were familiar with in their usual clinical practice. The data demonstrates that the age of ultrasound equipment in use by ultrasound practitioners has no effect on their subsequent quality rating of diagnostic ultrasound examinations. This should not therefore be a factor when determining which practitioners should undertake quality assurance.

### 5.3.4 Effect of Clinical Expertise.

There was no difference in scores between the 'peer-group' and 'expert-group' of study participants when looking at scores given for report quality and overall score. There was a small effect when looking at image quality with the expert group giving slightly higher scores than those of the peer group. The actual effect is small and

unlikely to be significant in the clinical setting. Because the expert group are more likely to have been in clinical practice for longer, this may be a residual effect from the result that more experienced practitioners rate image quality higher than their less experienced colleagues. Participant expertise and experience are not truly independent variables and it would therefore seem that a combination of these factors leads the more experienced (and more expert) ultrasound practitioners to give higher scores when assessing image quality.

#### **5.4 Is modelling possible in ultrasound quality assurance?**

One of the aims of this study was to determine whether there were systematic effects between participants. If so, it would be possible to model these effects to correct for individual participant characteristics. This is an attractive proposition as it would allow quality assurance to be undertaken by a wider and more diverse group of practitioners with appropriate correction factors put in place to account for this diversity. However, systematic differences between participants are not clearly defined and do not have linear effects on quality scores in this study. Application of such a model is therefore difficult given the limitations of this data.

The effect of clinical grade on report quality score is confusing. Participant grade has no effect on the scores given for image quality, but there is some effect on report quality score. Band 7 practitioners tend to rate ultrasound reports more highly than those in higher clinical bands. The reason for this may be due to the high level of interpretative skills required when reporting ultrasound examinations which may not be fully appreciated by those in lower clinical bands. Practitioners of lower clinical grade may be more willing to give (and accept) ultrasound reports which are

technically accurate but in which the interpretative element is lacking whereas participants of a higher clinical grade may deem these reports to be less acceptable. This is however, contradicted by the finding that there was no significant difference in scores given for the reporting element of this study between peer-group and expert-group of practitioners. The reasons why practitioners of a lower clinical grade score ultrasound reports higher than those in a higher grade is not therefore clear.

This data therefore shows some weak systematic effects on quality assurance scores given by study participants who were different in terms of clinical grade, expertise and experience. However, the data is not sufficiently robust to support implementing a theoretical model which would allow correction factors to be consistently applied to account for those differences in participant characteristics.

The data does demonstrate some interesting findings in the differences between study participants when assessing ultrasound examination quality. More experienced practitioners score image quality more highly than their less experienced colleagues. Conversely, in rating the quality of the written report, participants of a lower clinical grade give higher scores to written reports than those of a higher clinical grade. Expertise in diagnostic ultrasound based on national profile, lecturing and authorship seems to have no clinically significant effect on either assessment of image quality or report quality.

### **5.5 Who should undertake quality assurance?**

For this quantitative data to be meaningful in terms of practical application in individual ultrasound units, it is important to develop evidence-based conclusions about which practitioner groups are best placed to undertake quality assurance in medical ultrasound. The use of (nationally accepted) expertise has no effect on the scores given and this therefore should not be used as a factor when assessing the validity of a quality assurance review, particularly when this is at variance with a review from another, 'non-expert' practitioner. Those of greater experience appear to be more pragmatic in assessing image quality, probably relating to their experience in accepting the limitations of ultrasound in the 'difficult to image' patient. However, arguably the most important factor regarding quality of diagnostic ultrasound is the quality of the written report as this will be used by the referring clinician to guide clinical management of that patient. This appears to be affected by the clinical grade of those undertaking the reviews.

It is therefore recommended that those of a higher clinical grade (but not necessarily of greater length of experience) should be utilised to benchmark and set levels of acceptable practice in respect of quality of ultrasound reports. This does not preclude band 7 staff from undertaking quality assurance (peer review), but it is important that this process should be overseen and monitored by a senior practitioner due to their more robust expectations.

Peer review may be beneficial to the person undertaking the quality assurance in that it enables them to benchmark their own practice against their peers[85].

However, the primary purpose of a quality assurance program is to derive evidence

of acceptable performance and to provide data to drive internal quality improvements. Benefits to individual practitioners who undertake quality assurance tasks are an important, but secondary function. It is therefore recommended that while peer audit of performance is acceptable and may be beneficial, this process should be overseen by a senior practitioner to ensure that standards are defined, maintained and where possible improved.

## **Chapter 6**

### **Analysis of Participant Comments.**

## **Chapter 6.**

### **Analysis of Participant Comments.**

**This chapter outlines the strategy employed in analysis of free participant comments and the results of that analysis.**

#### **6.1 Introduction**

The website invited comments from participants on each of the sets of ultrasound imaging that they had reviewed. The comments section was a 'free-text' section, built into the website which was of unlimited length. There was no attempt to guide participants in terms of types or content of comments. It was completely up to each study participant as to whether they wished to comment and what form and substance those comments should take.

Because it was unclear as to how much participants would wish to comment and what form these comments might take, no structured procedure was put in place for how these comments would be analysed during the planning stage of this study. Most study participants chose to comment on the examinations that they had reviewed and this has provided an unexpected, but rich source of data. Although this project was designed as a quantitative assessment of current techniques in quality assurance of diagnostic ultrasound, the qualitative data is of great interest as it lends some context to the quantitative data. It was intended that this project be used to inform the development of robust quality assurance methods and mechanisms in non-obstetric diagnostic ultrasound. The quantity and detail of

qualitative data may help to inform how feedback should be given to clinical ultrasound practitioners to best inform continual improvements in clinical quality.

## **6.2 Analysis of Participant Comments**

Almost all reviewers chose to leave comments on the website regarding the ultrasound imaging that they had reviewed. The comments were of varying length, ranging from a few words to several paragraphs. Initial analysis of participant comments showed that comments could be subdivided into two broad categories; clinical ultrasound technique and clinical report writing.

### **6.2.1 Clinical Ultrasound Technique**

There were a wide range of comments on ultrasound technique. Comments ranged from technical settings used on the ultrasound machine, use (and accuracy) of anatomical labelling, clinical scope of the examination and overall quality of the ultrasound image obtained.

### **6.2.2 Clinical Report Writing**

Comments ranged from correct (or incorrect) use of grammar, perceived accuracy of the report, correct use of terminology, brevity of the report and perceived accuracy of any diagnosis given.

## **6.3 Analysis of Comments Strategy.**

In total, 176 sets of ultrasound examinations and accompanying reports were reviewed. Although not a requirement for inclusion in the study, participants were invited to leave comments regarding the ultrasound imaging that they had reviewed

and a mechanism was built into the project website to facilitate this. Most participants chose to leave comments, many of which were extensive.

Many of the comments were long and thus were subdivided into smaller units because they covered a wide scope of content on image and report quality. Thus, individual comments on either the image quality or clinical report quality aspects of the scan were treated as discrete comments, even if these were part of a longer discourse made by a study participant on a single case. This allowed the variety and richness of the comments to be appraised while keeping the volume of data at a manageable level. It explains why the numbers of comments analysed exceeds the total number of reviews undertaken (as many comments were divided into smaller discrete comments).

#### **6.4 Constructiveness of Comments.**

If this method of quality assurance is used to inform and drive quality improvement in diagnostic medical ultrasound, it is important that feedback and criticism are given in a constructive manner. Specific, constructive criticism maximises the probability of effecting improvements[86]. To assess whether feedback was constructive or non-constructive, objective criteria were necessary to determine the nature of the comments.

The work of Hameed and Mahmood[87] was used to inform this assessment. While some of the criteria specified by these authors are not appropriate for internet-based

feedback (e.g. feedback should be opportunistic and/or well-timed), most can be successfully modified for internet based, rather than direct, face-to-face appraisal.

Feedback was deemed constructive if it met the following criteria;

- **Balanced:** appreciation of the good and bad
- **Clear:** in terms of criterion and applicable standards.
- **Encouraging:** for time and effort.
- **Factual:** based on actual performance rather than on assumptions or interpretations.
- **Focused and Specific:** Focusing on the observable and changeable elements of performance.

To assess whether criticism was constructive or non-constructive, the comments for each of the 176 study reviews were assessed in their entirety. Many of the longer comments contained both constructive and non-constructive elements. In this case, a decision was taken by the individuals assessing the comments as to whether the overall tone and content of the comments was either constructive or non-constructive. Comments were assessed in a blinded fashion to reduce the probability of bias during analysis of these comments. Participant identification was coded so that the individuals analysing the comments were not aware of the background or individual characteristics of individual study participants during the analysis of their comments. Analysis of comments was undertaken independently by two assessors. One had specialist knowledge of diagnostic ultrasound. The other assessor was a registered senior nurse with no specialist knowledge of ultrasound. The use of a non-specialist was considered important to try and disentangle the

effects of tone and language of comments from the technical contents of those comments.

#### 6.4.1 Constructive Comments.

A typical constructive feedback for a high-quality ultrasound examination is given in Figure 10 as an example;

*'High quality exam. High quality images and report. My only minor change would have been to reduce the description of the calculus in the report. I doubt twinkling sign would mean much to oncologists or urologists and is, arguably, unnecessary jargon. Otherwise, great.'*

Figure 4. Constructive comment for high quality ultrasound examination.

Constructive criticism is perhaps more important for ultrasound examinations of lower clinical quality. Even where feedback was critical, some study participants still managed to make comments constructively, even with studies of perceived lower quality, as demonstrated in Figure 11;

*'Portal hypertension from diffuse liver disease is still a possibility, despite a normal looking liver. A high frequency probe to look at the liver capsule and parenchyma would have been useful in either including or excluding this as a possibility. The cause of ascites is not established - perhaps a look at the pelvic area and bowel might have been more comprehensive.*

*No advice in the report as to subsequent management.*

*Images OK, but could improve the image optimisation in a number of cases - they all seem to have been done with the same depth, sector angle and focal zone position'*

Figure 5. Constructive feedback for ultrasound examination of lower quality.

#### 6.4.2 Non-Constructive Comments.

Some comments focused entirely on the negative aspects of the ultrasound examination and report without always giving specific information as to what the problems were or how the examination or report could have been improved, an example of which is given in Figure 12.

*'No optimisation of image. No annotation. Multiple images of the same organ - all poor quality. Poor portal vein imaging - not confirming normality. No imaging of hepatic veins. No mention of intra or extra hepatic ducts. Poor images of the pancreas – can't confirm normality. No mention of ascites or not. Bladder mentioned but not imaged.*

*Borderline splenic enlargement - not referred to.'*

Figure 6. Comment judged as Non-constructive by both assessors.

Some study participants chose to simply re-report the ultrasound examination based on the static imaging available to them. This was classed as non-constructive feedback as it may not be factual, nor was it focused, nor specific. An example is given in Figure 13.

*Cystic mass arising from the pelvis thought to be ovarian in origin. mural wall thickening seen no Doppler flow. DDx mucinous cystadenoma correlation with CA125 level. Urgent CT Abdomen pelvis and Gynaecology oncology referral.*

Figure 7. Non-constructive feedback, re-reporting the ultrasound examination.

## **6.5 Methodology of Comment Analysis.**

The comments were analysed according to the following variables.

### **6.5.1 Comment content.**

Numbers of comments related to ultrasound clinical technique were compared with the number of comments relating to the quality of clinical report writing. The rationale for this was to examine whether reviewers were more concerned with report quality, image quality or a combination of both.

### **6.5.2 Comparison of comments from the 'expert-group' compared with the 'peer-group' of participants.**

Comment content was compared from participants belonging to the peer-group with those belonging to the expert-group. The rationale for this was to try and understand whether reviewers of differing expertise looked at the ultrasound images and reports differently. It was important to try and establish whether reviewers of differing levels of expertise were more concerned with the quality of the ultrasound image, the quality of the clinical report or a combination of the two.

### **6.5.3 Comparison of constructive and non-constructive reports as judged by a specialist in diagnostic ultrasound.**

An assessment was made as to whether the comments were constructive or non-constructive in nature, and whether there were systematic differences between study participants of different expertise levels. Content and tone of the comments is clearly very important in terms of providing encouragement to individual practitioners and incentivising quality improvements. It was felt important to try and establish

whether there were any differences between the expert and non-expert group of participants in terms of the number of constructive and non-constructive comments given.

#### 6.5.4 Comparison of constructive and non-constructive reports as judged by a non-specialist in diagnostic ultrasound.

During analysis of comments by the specialist in diagnostic ultrasound, the difficulties in disentangling technical details from the overall nature and tone of the comments became apparent. This was a potential source of bias as the ultrasound specialist may have been swayed by some of the more technical aspects of the comments while not evaluating overall tone and content adequately. It was therefore decided that a non-specialist senior healthcare practitioner, with no specialist knowledge of ultrasound would be invited to review the comments for tone and language alone, and decide whether they were constructive or non-constructive in nature. The individual approached for this role was a senior nurse specialist who had extensive experience in both training, mentoring and assessing trainees and junior members of staff.

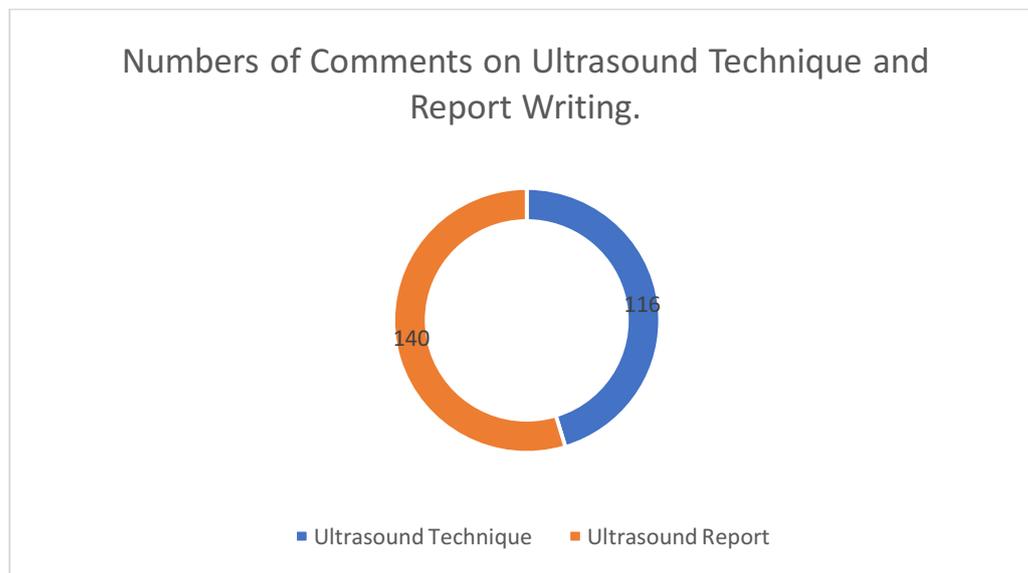
### **6.6 Statistical Analysis.**

The null hypothesis was that there would be no significant difference between expert and non-expert study participants in terms of the content and constructiveness of their comments. Chi squared test was used to test this hypothesis with rejection of the null hypothesis at a p value of 0.05 or below.

## **6.7 Results.**

### **6.7.1 Content of Comments.**

In total, 256 individual comments were reviewed and assessed from 176 participant reviews. The number of comments regarding quality of the clinical report compared with the quality of the ultrasound imaging itself were approximately equal. 55% of comments concerned the quality of the clinical report and 45% were concerned with clinical technique. This is demonstrated graphically in Figure 14.



**Figure 14. Comment Content for all participants.**

### **6.7.2 Analysis of comments by expertise of participant.**

Table 22 demonstrates the content of comments given against the expertise of the study participant. There was no significant difference in content of comments between the expert-group and peer-group of study participants ( $p=0.48$ ). The level of

expertise of the participant had no effect on whether the comments emphasised clinical technique or the clinical report quality.

| <b>Table 22</b>   |                              |                            |       |
|---|------------------------------|----------------------------|-------|
| <b><i>Comment number by Subject and Participant Expertise</i></b> |                              |                            |       |
|   | Expert-Group<br>Participants | Peer-Group<br>Participants | Total |
| <b>Comments regarding<br/>Image Quality</b>                       | 59                           | 57                         | 116   |
| <b>Comments regarding<br/>Report Quality.</b>                     | 65                           | 75                         | 140   |
| <b>Total</b>  | 124                          | 132                        | 256   |

Table 22. Comments by Subject and Level of Participant Expertise.

## **6.8 Constructiveness of Comments**

### **6.8.1 Clinical Ultrasound Specialist Observations.**

Using the criteria previously described, the comments for each ultrasound examination review were assessed by a specialist in medical ultrasound to determine whether the overall content and tone was constructive or non-constructive. Where the comments had both constructive and non-constructive elements, a judgement was made as to whether the feedback in its entirety was either constructive or non-constructive.

116 comments were assessed in total. There were slightly more constructive than non-constructive comments with 59.5% judged as constructive and 40.5% comments judged as non-constructive. The results are shown in Figure 15.

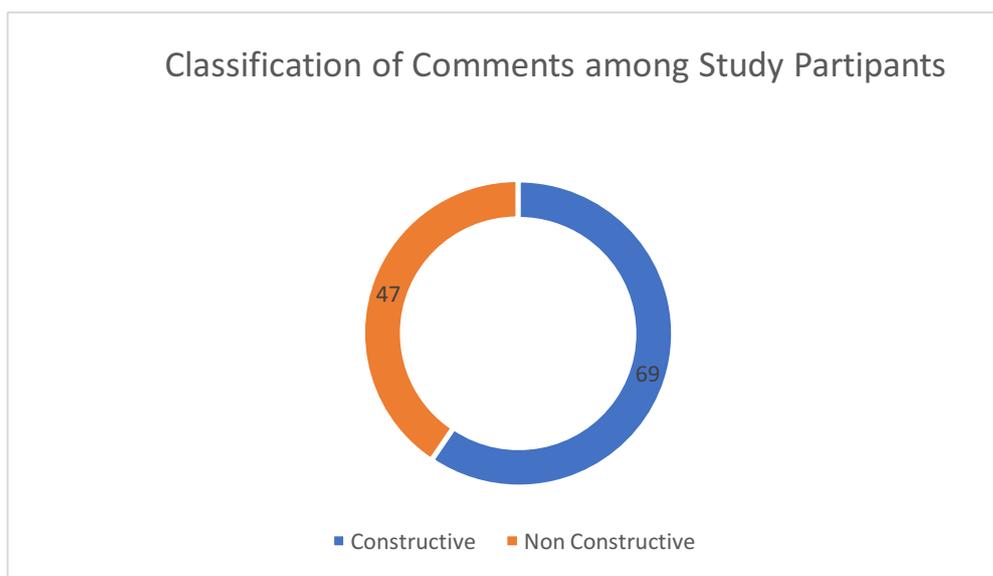


Figure 15. Constructiveness of Comments among Study Participants.

### 6.8.2 Classification of Comments according to Participant Expertise.

The number of constructive and non-constructive comments were analysed according to the expertise of study participants (Table 23)

Table 23.

***Constructiveness of Comments per Participant Expertise as judged by an Ultrasound Specialist.***

|   | <b>Expert-Group Participants</b> | <b>Peer-Group Participants.</b> | <b>Total</b> |
|---|----------------------------------|---------------------------------|--------------|
| <b>Number of Constructive Comments</b>      | 53 (76.8%)                       | 16 (23.2%)                      | 69 (59.5%)   |
| <b>Number of Non-Constructive Comments.</b> | 12 (25.5%)                       | 35 (74.5%)                      | 47 (40.5%)   |
| <b>Total</b>                                | 65 (56%)                         | 51 (44%)                        | 116          |

Table 23. Constructiveness of comments per participant expertise as judged by an ultrasound specialist. p<0.01

In the view of the ultrasound specialist, the expert group were more likely to offer constructive criticism than the peer group. 81.5% of the expert group's comments were assessed as constructive. In contrast, the comments offered by the peer group were classed as constructive in only 31.3% of cases. The differences in number of constructive vs non-constructive comments between the expert and peer group are clinically and statistically significant with a  $p < 0.01$ .

### 6.8.3 Non-Ultrasound Specialist Observations.

The non-specialist assessor did not differ significantly from the specialist assessor in classifying the total number of comments as constructive or non-constructive.

Taking the participant group as a whole, there were slightly more constructive than non-constructive comments with 58.6% comments judged as constructive and the remaining 41.4% judged as non-constructive (Figure 16).

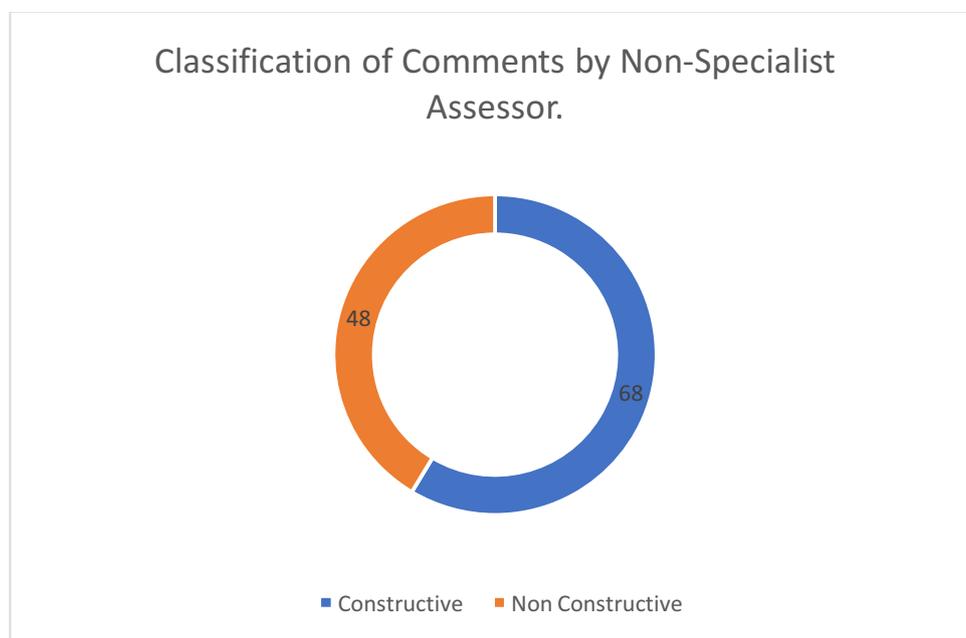


Figure 16. Constructiveness of comments as judged by a non-specialist observer.

### 6.8.5 Classification of Comments by Non-Specialist Assessor.

There was a marked difference in the way that the non-specialist assessor judged the comments compared with the ultrasound specialist assessor. Although the general trend for expert participants to give more constructive comments than non-expert participants continued, this effect was not nearly as marked when the comments were reviewed by a non-specialist observer.

In the view of the non-specialist assessor, comments made by the expert-group of participants were judged to be constructive in 60% of cases. Comments made by the peer-group were judged to be constructive in 46% of cases. These results are not statistically significant with a p value of 0.48. When judged by an independent, non-ultrasound specialist, there is therefore no significant difference in constructiveness of comments between expert and non-expert participant groups. This data is summarised in Table 24.

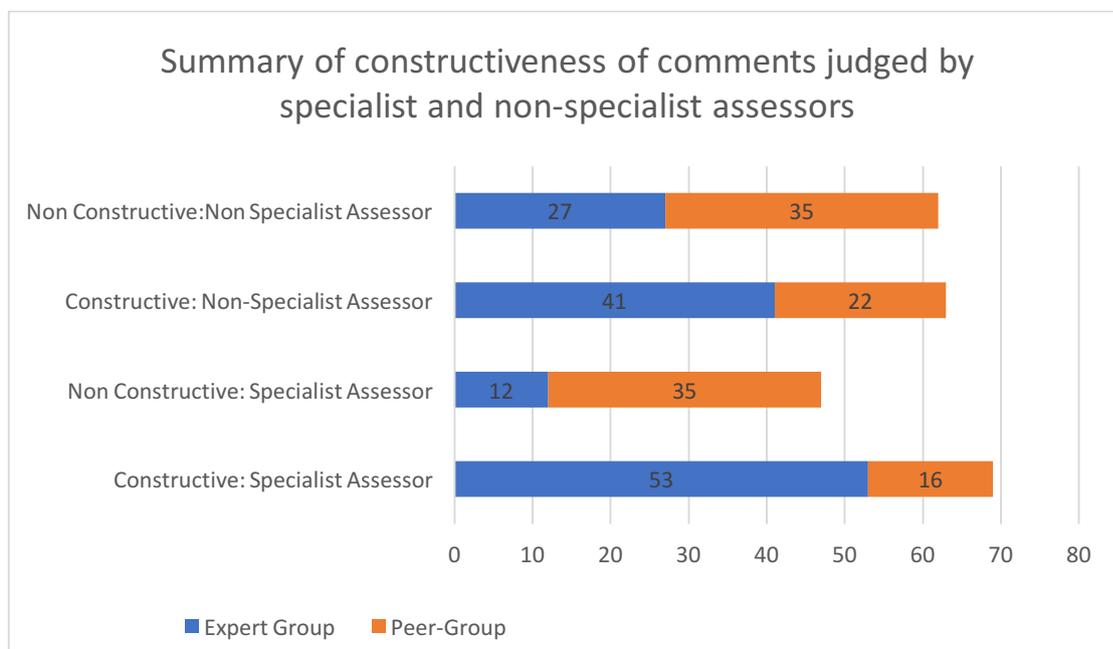
| Table 24<br><i>Constructive vs Non-Constructive Comments judges by Non-Ultrasound Specialist</i> |                                      |                                     |              |
|--|--------------------------------------|-------------------------------------|--------------|
|  | <b>Expert Group<br/>Participants</b> | <b>Peer Group<br/>Participants.</b> | <b>Total</b> |
| Number of Constructive Comments  | 41                                   | 22                                  | 63           |
| Number of Non-Constructive Comments.   | 27                                   | 26                                  | 53           |
| Total  | 68                                   | 48                                  | 116          |

Table 24. Constructive vs non-constructive comments as judged by a non-ultrasound specialist. p=0.48

## **6.9 Differences between specialist and non-specialist assessors in comment analysis.**

In assessing the differences in comment analysis between expert and non-expert study participants, whether the individual analysing the comments is a specialist or non-specialist in this field is clearly of importance. The differences between the specialist and non-specialist assessor are summarised in Figure 17.

When judged by a specialist ultrasound practitioner, there is a significant difference between expert-group and peer-group participants with comments being perceived as more constructive when made by expert participants. However, this effect was not replicated when the comments were reviewed and analysed by a non-specialist practitioner. The potential reasons for, and implications of these results will be explored in Chapter 7.



**Figure 17. Comparison of Specialist and Non-Specialist Comments Analysis**

## **Chapter 7**

### **Discussion of Participant Comments.**

## **Chapter 7.**

### **Discussion of Free Comments made by Participants.**

**This chapter aims to examine in greater depth the analysis of comments made by study participants. This chapter intends to form the basis of recommendations around giving qualitative feedback in ultrasound quality assurance.**

#### **7.1 Introduction**

During the design phase of the website, it was felt important to give participants the freedom to make comments on the ultrasound imaging that they had reviewed if they so wished. The principal researcher was mindful of the extensive time required for research participants to undertake the reviews and was therefore wary about introducing an additional burden on study participants by placing them under obligation to leave comments when they may not have wished to. It was therefore decided to design an unstructured, free-comments box into the website where participants could leave comments if they wished, but no constraints or guidance was offered on what form or length that these comments should take or indeed whether any comments should be left at all.

The comments section of the website has provided a very rich vein of data due to the willingness of study participants to leave extensive comments on their opinions regarding the imaging that they had reviewed. The number, detail and depth of the comments was unexpected and a comprehensive plan for qualitative data analysis of comments was not therefore factored into the original study design. However, the

number of comments and nature of feedback has offered the opportunity for a basic level of analysis, the results of which have been described in Chapter 6. This chapter attempts to interpret these results by placing them within the context of a theoretical background of mechanisms of giving feedback and the potential effects of those mechanisms. It hints at how the study design may have affected how feedback was both offered and received. Finally, it signposts the need for further qualitative study in this area and how training for those offering feedback may assist in ensuring a positive outcome from a feedback intervention.

## **7.2 Analysis of comments.**

### 7.2.1 Comment content.

Analysis of the comments demonstrated roughly equal proportions of comments regarding the quality of images and comments regarding the quality of the report intended for the referring clinician. This suggests that when assessing an ultrasound examination for quality, study participants place roughly equal weight between the quality of the ultrasound images and the quality of the clinical report based on those images. This was true for participants in both expert and peer groups of the study. Level of expertise therefore appears to have little effect on what study participants look for when reviewing ultrasound images and clinical reports with the review tools tested within this study.

### Differences between 'Peer' and "Expert" participants.

There was an interesting difference between the expert-group and peer-group of participants in terms of the perceived constructiveness of the comments offered. When comments were analysed by a specialist in diagnostic ultrasound, there was a statistically and clinically significant tendency for comments made by the expert

study participants to be rated as more constructive than those offered by peer-group participants. However, this result was not confirmed when participant comments were analysed for language and tone by a non-specialist in diagnostic ultrasound. Analysis of participant comments was undertaken in a blinded fashion. Neither individual undertaking comments analysis knew which comments originated from expert or peer group participants. Both individuals undertaking the comments analysis were also blinded to the other's results. These results suggest that the technical detail and content of the comments were more important than tone to the specialist in ultrasound. Conversely, the non-specialist in ultrasound was more likely to be interested in the tone and language used within the comments.

To fully understand and unify these conflicting results, it becomes necessary to find a theoretical model that provides a backdrop against which these results can be understood. It was important to try and understand the importance of the mechanism of a feedback intervention and how the method of offering such feedback may affect the outcome of such feedback, both positively or negatively. By analysing these results and placing them within this theoretical context, a better understanding can be gained regarding the best conditions under which feedback should be given.

It should be remembered that study participants were not given any guidance in leaving their individual comments. They were also not aware that any comments they made would be used to try and extrapolate data regarding how best to give feedback to ultrasound practitioners. If this had been known by study participants, it

could be argued that their comments may have been different. This data must therefore be treated with some caution, yet it does provide a fascinating insight into models and mechanisms of offering feedback and opens avenues of areas of potential further research.

### **7.3 Use of anonymous feedback.**

Just over 40% of comments were judged as non-constructive when judged against well accepted, specific criteria by both a specialist and non-specialist reviewer. The reasons for this require exploration. It may be due to a lack of training in offering constructive feedback, a lack of experience or a lack of emotional intelligence on the part of the study participants.

It is unclear whether the anonymous nature of the study methodology which enabled participants to safely offer anonymous criticism online, influenced the quality of qualitative feedback given. There is little literature comparing the quality and tone of feedback when offered face-to-face compared with feedback given anonymously over the internet, and none in this specialist field.

Because of the lack of primary research in the field of diagnostic ultrasound, it is valid to draw parallels from similar, analogous subject areas where there is more data and knowledge. Both peer review in the medical literature and the process of clinical assessment of trainees have some similarities with this study. Peer review of the medical literature generally involves reviewers reducing assessment of a scientific paper into objective and subjective comments and criticism. During trainee assessment, feedback may be given face-to face or online and review of this literature is also helpful in interpretation of the results of this study.

This study was not designed to assess the effect of anonymised compared with non-anonymised reviews, but if peer review over the internet is to be a serious option in quality assurance in ultrasound, the effect of this methodology does warrant consideration. The effect of anonymisation in offering feedback has been researched in both peer assessment in education and peer review for publication. Some indicate that reviewers are more likely to offer higher quality reviews if anonymous[81, 88] while others found no effect[79, 80, 89]. It has been suggested that reviewers are more likely to offer open and more stringent criticism when feedback is given anonymously[88]. This may be particularly important if ultrasound quality assurance is undertaken as a peer review exercise as ‘anonymity protects younger, less powerful reviewers from possible retribution’[90].

Li et al[85] found that when students were asked to peer review other student’s projects, the quality of reviews given was more predictive of the reviewer’s (rather than reviewee’s) ability. Although the context is very different, it therefore seems that the more able practitioners are more likely to give higher quality reviews of ultrasound imaging. Because more able practitioners are more likely to be regarded as ‘expert’ within the ultrasound community, this may explain why the expert-group of participants were perceived to offer more constructive comments in the eyes of the specialist reviewer, even without formal training.

## **7.4 Qualitative feedback mechanisms**

### **7.4.1 Purpose of feedback.**

The purpose of offering feedback is to improve professional performance, changing ‘provider behaviour and consequently quality of health care’[91]. There is the natural

assumption that feedback is an intervention which universally improves outcomes, yet there is evidence that clumsily undertaken or poorly-planned feedback strategies can actually worsen professional performance in up to one third of cases[92]. It is therefore imperative that provision of feedback to individual practitioners is performed carefully, thoughtfully and tactfully. To achieve this, it becomes necessary to examine mechanisms of offering feedback in detail. This understanding should assist in creating a culture where feedback is offered in a positive manner which positively affects performance.

#### 7.4.2 Theories of feedback

To underpin and guide the practice of giving constructive feedback, it is helpful to examine some of the psychological theories which explain this area of practice. Kluger and DeNisi[92] offer a feedback intervention theory which comprises the following principles.

- **Behaviour is regulated by comparisons of feedback to goals or standards.**

It would seem self-evident that for feedback to be meaningful, it needs to be applied against a set standard. One of the major findings from this study is that the application of meaningful standards in the assessment of diagnostic ultrasound in the non-obstetric setting is challenging. The quantitative data demonstrates only 'fair' agreement between study participants when assessing what constitutes a 'good' ultrasound examination and accompanying clinical report. The level of agreement between participants is unaffected by the level of expertise of the study participants. This strongly suggests that there are currently no such applicable standards which can be consistently applied between reviewers.

- **Cognitive attention is limited and therefore only feedback-standard gaps that receive attention actively participate in behaviour regulation.**

Feedback-standard gaps refer to the gap between actual performance and the expected performance as set by the standard. Thus, this principle refers to the fact that behavioural change can only be expected when attention is drawn to the gap between measured and expected performance, either negative or positive.

- **Attention is normally directed to a moderate level of hierarchy.**

Hierarchy in this context refers to Action Identity Theory[93]. This theory describes how different levels of meaning can be applied by different individuals to the same act. Actions low in an individual's hierarchy are those performed without conscious effort or thought. They may comprise a series of detailed components which make up a single act. An action which lies high in an individual's hierarchy is more closely related to the sense of self and ultimate goals of the self. They often comprise significant thought and effort. The position of an action within an individual's hierarchy will vary from person to person and may also change over time in the same individual. Feedback interventions are most effective when aimed above the level of very detailed individual components of the task, often performed without conscious effort, yet not pitched to a level where they affect the ultimate goals of the self[92].

- **Feedback interventions change the locus of attention and therefore affect behaviour.**

Feedback interventions will usually have implications on the 'sense of self' of the person to whom feedback is being given. Interventions are therefore very unlikely to be ignored. The way that feedback is given will affect the reception at different levels within the psychological hierarchy and is therefore of great importance.

By utilising a theoretical model of giving feedback, the effect on subsequent behaviours of the recipient can be more clearly understood. The mechanism of giving feedback is clearly very important in determining whether it exerts a positive, neutral or negative effect on subsequent actions and behaviour.

#### 7.4.3 Methods of giving feedback and their effects.

There has been considerable interest and research into the effects of giving feedback in medical education, (particularly undergraduate education) since the 1980s[94]. While this may not be directly applicable to providing feedback for qualified and experienced ultrasound practitioners, the literature does serve as a useful guide when planning for, and giving feedback, particularly when this is negative in nature. The key aspects are summarised as follows[95];

- **Work as an ally of the person receiving feedback.**

Time and place should be agreed between those giving and receiving feedback. The giver of feedback should encourage those receiving feedback to offer a self-assessment of their own performance before offering their own opinion. Any

feedback should be offered against well-defined and mutually agreed outcomes.

- **Feedback should be based on observed and modifiable performance and behaviours.**

Feedback based on subjective data and interpretations should be stated as such. The person giving feedback should ensure that feedback is objective as far as possible and based on what is observed. Attempting to interpret the attitudes and motivation behind an element of performance should be avoided. Feedback should be given on specific behaviours rather than general performance. Feedback should be given in small quantities to avoid overwhelming the recipient of such feedback. Language should be carefully chosen to avoid being judgmental.

### **7.5 Application of feedback guidance.**

These factors can be used in terms of assessing how effectively feedback can be applied under the constraints of this study. At first glance, many of these factors are not possible to achieve when using internet-based feedback methods. Although the receiver of feedback can choose a time to receive the feedback (being free to log onto the website at a time convenient to them), there is little scope for meaningful discussion under this system. The receiver is merely picking up feedback left previously by a reviewer of their work and performance. In terms of basing feedback against mutually agreed outcomes and standards, the quantitative data suggests that there is no current agreement within the ultrasound community in terms of acceptable standards in quality of ultrasound images and reports.

Attempting to generate guidance on how feedback may be offered under these conditions is therefore challenging. Even given these limitations, some of the comments left by study participants could be construed as judgemental and lacking in specific advice on how to improve performance. Use of language sometimes appeared to be ill-judged and unnecessarily negative. Feedback of this type was thought unlikely to have a positive impact on subsequent performance if it were to be given to an individual practitioner. An example is given in Figure 18. Please note that this comment is in 'raw' format and has not been edited to correct spelling or grammar.

*'Poor image optimisation. Little attempt to image the liver with altered TGC or frequency to penetrate.*

*Looks like poor technique too - repetitive (poor) images of the same area - e.g. 5 identical pancreas images contributing nothing, 6 almost identical gb images - no transverse - none of the neck and no attempt to turn the patient or image from a different window - e.g. intercostally.*

*Looks like the patient has not really been moved onto his/her side and no real attempts at intercostal scanning . Segments 7 and 8 have not really been demonstrated here.*

*Spleen poorly measured, and 13.9.cms is, in any case, enlarged not normal as stated.*

*We never quite see the upper pole of the RK - no attempt at intercostal scanning to demonstrate this or of a posterior approach.*

*No labels - so I'm assuming it's the RK!*

*Report writing contains some shockers! 'Bright' should never appear in a report. The phrase 'normal echo appearances' should be banned.*

*The statement that no focal lesions are seen is misleading as the operator has not produced adequate comprehensive images of the liver. It is fatty - correct - but not 'coarse' as stated.*

*This has shades of a Y1 stumbling through a scan with no real idea?'*

Figure 8. Negative Comments made by a study participant.

There is a large amount of primary research into the best ways of giving feedback to elicit a positive response. A large Cochrane systematic review [96] analysed 149 studies which had a significant feedback-giving component and noted that the actual processes involved in giving feedback have a small but significant positive effect on improvement overall. The margin of improvement is estimated at 4% although the variation was wide between studies.

In general, the Cochrane review largely mirrored the work of Ende[94] in that feedback was most effective when given more than once, both verbally and in writing and included clear targets with an action plan stating how those targets were to be achieved. Feedback was most effective when the person offering the feedback was either a colleague or supervisor, known to the individual receiving feedback.

Baseline performance of a practitioner also affects how much impact a feedback intervention has. Feedback is most effective when offered to those individuals who have a low initial baseline. i.e. those who are already performing poorly when the feedback is offered. This fits well with feedback intervention theory which suggests that where there is a large negative gap between perceived performance and external standards, the desire to improve performance moves up the psychological hierarchy with more incentive and motivation to improve performance. These factors have been validated when applied to radiographers and student radiographers[97]

### **7.6 Feedback guidance and study conditions.**

For the purposes of this study, it was important to assess how well the results of the Cochrane systematic review could be applied to the conditions under which this study was performed. There were significant difficulties in implementing the

recommendations for offering feedback within this study. The conditions imposed by internet feedback, mean that feedback cannot easily be given verbally. In addition, utilisation of an external, (possibly) unknown reviewer makes reception of such feedback more difficult because the person providing feedback is not a direct colleague or supervisor.

While some improvements may be possible in practitioners who are not performing optimally due a large negative feedback-performance gap, for most practitioners who are performing either adequately or well, feedback will have less of an effect due to the small or non-existent feedback-standard gap. Viewed in this context, the assumption that anonymous written feedback will have the effect of making small positive changes in practice among already high performing practitioners as part of a continuous quality improvement program is rather optimistic. The conditions imposed by this study are not optimal for offering constructive feedback which can be accepted and acted upon to produce positive subsequent improvements.

There remains little literature around the effect of giving feedback via the internet on subsequent objective, measurable changes in performance. The measurable effect of an on-line feedback intervention remains frustratingly small among those studies undertaken[98-102]. However, while accepting that the internet does not provide optimal conditions for feedback, it remains important to try and establish whether improvements can be achieved to enhance the experience of feedback for givers and receivers alike, within the constraints of web-based feedback.

### **7.7 Training in Giving Feedback.**

The literature suggests that conditions imposed by this study were not optimal for either providing or receiving feedback which could positively affect change in practice. This study elicited many non-constructive comments, as judged by both a specialist and non-specialist in ultrasound. Approximately half of the comments were classified as non-constructive which suggests that the negativity of these comments cannot be attributed exclusively either to the technical content, or tone and language of these comments. The ultrasound specialist felt that the expert-group offered more constructive comments than the peer-group of study participants. This suggests that technical content plays a significant part in how these comments may be received. However, the non-specialist assessor concluded that the tone and language of the comments remain similar between expert and peer group study participants.

It is unknown how many of the study participants had received specific training in giving feedback. However, it was clear that study participants were unable to offer constructive feedback in a significant proportion of cases. It should be borne in mind that the study participants were offering comments without the direct knowledge that these would be analysed specifically as feedback and some caution should therefore be used in the interpretation of these results. However, in approximately half of cases, the negative aspects of an individual ultrasound examination and report were expressed in such a manner that this was deemed to be non-constructive by both a specialist in diagnostic ultrasound and non-specialist health care worker.

Although the internet is not an ideal forum for giving or receiving feedback, there is clear potential for improvements to be made which *are* possible within the constraints

of the study methodology. Training in provision of feedback may be helpful in improving the quality of the feedback, maximising the odds that the feedback will be translated into positive change.

The existing evidence is mixed regarding the effect of formal training in offering feedback. Among surgeons, there is evidence [103] that around 50% of surgeons significantly overestimate their performance in giving effective feedback, even among those with a degree in medical education. While it is doubtful that this can be directly translated to ultrasound practitioners, it does imply that formal education does not automatically enhance the quality of feedback given when compared to those without a formal qualification in education. There is a larger body of evidence which suggests that offering training in giving feedback does improve the outcomes from that feedback[97, 104-109]. While much of this evidence relates to trainees or newly qualified staff (rather than established practitioners), the evidence suggests that targeted training in offering feedback is effective. This is the case over a wide range of subjects in both health care and non-health-care disciplines. This has also been examined in the specialist fields of general radiography and ultrasonography.

### **7.8 Who should provide feedback?**

The evidence is that feedback should be given by an individual well-known to the receiver of the feedback for it to be effective[94, 96, 110]. Unfortunately, this does not sit comfortably with the methodology of this study in that the data from those offering feedback was anonymised and therefore unknown to those who had personally undertaken the ultrasound examination. It is unlikely that the feedback

offered would be well received by the recipient for this reason, and therefore the likelihood of facilitating positive change is reduced. In addition, given that a significant proportion of the feedback was also assessed as non-constructive, if this feedback were given to ultrasound practitioners under the conditions of this study, there would be significant risk of a negative outcome from such feedback, related to dismissive or defensive reactions by the recipient[111].

Whether feedback should be offered by peers or superiors is also an important question. Peer review has been trialled in ultrasound with some encouraging results although this method of review has yet to become well established[112]. Peer feedback offers a practitioner the opportunity to assess their own performance against established group norms. This can be a powerful motivator in changing behaviours which do not conform to those norms[113].

The potential disadvantage of introducing a review scheme based on peer-feedback is the risk of establishing group norms which are below those that would be expected by an external, independent or expert observer[39]. The risk is that the establishment of such group norms may result in poor department culture and performance as judged by an external observer[39]. In addition, the validity of such reviews may be in doubt as there is some evidence that peer review is more predictive of the reviewer's (rather than reviewee's) ability[100].

While peer feedback may have little impact on the performance of others, there is evidence that it does have a significant impact on performance of the person giving

feedback[85, 114]. By having the opportunity to give feedback, practitioners have the opportunity for self-reflection on their own performance, compared with the performance of their peers, resulting in improved performance. This could not be tested under the methodology of this study, but the effect of giving peer feedback on performance in radiography and particularly ultrasound would seem a fertile area for further investigation.

#### 7.8.1 Benefits and difficulties of feedback

The giving of honest, yet constructive feedback can be difficult and may be avoided due to difficulty in successfully addressing the issue. There appears to be a reluctance, particularly among doctors to offer criticism of professionally or ethically questionable behaviours, yet those who are able to offer criticism are found to receive more professional support from colleagues when being criticised themselves[115]. While it remains difficult to establish whether peer feedback has a significant impact on maintenance of safe working practices and standards at a local level, the evidence suggests that creating a culture that empowers practitioners to offer peer feedback confers definite advantages on those that offer that feedback. It serves to both improve their own practice and places them in a position in which to receive support from colleagues themselves when things go wrong.

Having feedback from a practitioner of a higher grade is also useful, as that practitioner is in a position of authority to both establish standards of practice and ensure that these standards are adhered to. These standards may be less stringent

or effective if decided entirely among a peer group of practitioners without anyone being in a position of authority to make a final decision on the boundaries of acceptability. Feedback is most successful when made against accepted and agreed standards and when targets and an action plan can be agreed. [94, 96]. Agreeing targets and an action plan is the responsibility of a supervisor who has line management responsibility over an individual practitioner, and for this reason a supervisor may be in the best position to offer feedback to an individual practitioner regarding their performance.

## **7.9 Conclusion**

While this study was not specifically designed to assess the effectiveness in offering feedback to ultrasound practitioners in their work, the feedback left by study participants has offered some tantalising glimpses into the practice of offering feedback to established practitioners in diagnostic ultrasound. Utilisation of a psychological framework is useful when deciding how, and by whom feedback should be given as it offers a context against which the mechanisms for feedback can be designed to optimise the chances of a positive outcome. Clearly, the mechanism, tone and content of feedback is of crucial importance. Improperly designed mechanisms or clumsily undertaken feedback interventions have the potential to precipitate a negative response in day-to-day professional performance and clinical practice.

Use of impersonal feedback via the internet is far from ideal, yet utilising the internet to provide regular, external appraisals of individual and department performance

probably confers benefits that outweigh the difficulties in offering on-line feedback. These include the advantage of utilising a detached and external reviewer which can help to benchmark clinical practices at a regional or national, rather than a local level. This wider system would seem more effective in establishing and maintaining effective standards of practice. However, ensuring that feedback has a positive (rather than neutral or negative) effect is clearly a challenge.

It is critical that written feedback conforms to a well-researched and methodologically sound framework to maximise the chances of this creating positive change. While there appear to be differences in the way feedback is perceived by specialist and non-specialist reviewers, many of the comments in this study were found to be non-constructive in either tone or content. This is of concern if these comments were to be used as part of a quality improvement program within an ultrasound department.

At present, there is evidence that a significant number of expert and peer group ultrasound practitioners lack the theoretical grounding and practical skills to offer effective feedback which would lead to positive outcomes in standards and performance. Coupled with the disadvantage of providing this feedback via the internet, the methodology used in this study is not optimal in providing feedback in such a manner to effect positive changes. However, outcomes could probably be improved by provision of formal theoretical and practical training in offering feedback designed to enhance the probability of positive outcomes in performance.

The data from this study raises the possibility that the current method of offering qualitative feedback may do more harm than good. There is scope for further research into this area, particularly investigation into how introduction of a training program in offering written feedback may affect the quality of the feedback offered. One of the outcome measures of such a study should be whether such a training program ultimately improves standards within diagnostic ultrasound departments.

Supervisors and line-managers are best placed to offer interventions in terms of target setting and action plans, particularly for poorly performing practitioners. However, there is scope for peers to also review the work of others. Evidence suggests that this is a good exercise in terms of continuing professional development because it allows the person offering feedback to benchmark the quality of their own practice against that of their peers. While there is little evidence to suggest that peer-to-peer feedback is an effective strategy in improving the practice of others, it does appear to enhance the practice of individual practitioners. Utilisation of peer reviewers could therefore raise the overall standard of a department through a process of self-development. There is the additional spin-off benefit that by enabling constructive feedback and criticism to be given between peers, individual practitioners are more likely to be offered support by their peers when errors are made.

## **Chapter 8.**

### **Recommendations for Clinical Practice.**

## **Chapter 8.**

### **Recommendations for Clinical Practice and Further Research.**

**This chapter gives evidence based recommendations for clinical practice based on the data from this study.**

#### **8.1 Introduction.**

This chapter seeks to summarise those aspects of this research project which have a direct impact on current clinical practice and aims to make evidence-based recommendations based on the data collected. Issues highlighted within this study centre mainly around levels of inter-rater agreement, systematic differences between groups, use of the internet as a mechanism for delivering and receiving feedback and the ability of study participants to give feedback in a manner which are most likely to result in subsequent positive outcomes and performance.

#### **8.2 Levels of inter-rater agreement.**

The level of agreement between study participants was only 'fair' raising the question as to whether this method of quality assurance in diagnostic ultrasound is of sufficient efficacy to put into routine, clinical practice. Because, there is a paucity of primary research in this specialist subject area, it is difficult to find a benchmark against which to appraise these results and this is therefore a difficult question to answer.

By researching a similar, analogous subject (peer review in the scientific literature), it has been possible to gain some impression on what level of inter-rater agreement variation is acceptable and whether a degree of variation between study participants may be beneficial in terms of quality improvement and advancing clinical practice.

Peer review of the scientific literature is a process that also experiences relatively high levels of inter-rater variation which is difficult to control for, even after introducing interventions to reduce this through training and education. However, the scientific peer review process succeeds in managing this variation by both utilising multiple reviewers and ensuring varying degrees of anonymity to ensure that reviews are both given and received in a fair and constructive manner.

There is anecdotal evidence that some ultrasound service providers are utilising a single 'expert' reviewer to review the work of ultrasound practitioners, particularly in evidencing contractual compliance to quality standards[73]. Given that levels of inter-rater agreement between reviewers is only 'fair' it is likely that the quality assurance undertaken by a single quality assurance practitioner is as likely to give as much information on the standards of the reviewing practitioner as on the ultrasound studies under review.

The use of a single reviewer (no matter how expert) should therefore be resisted. If this mechanism of quality assurance is to be used effectively, multiple reviewers should be utilised to buffer against the effects of high levels of inter-rater variation.

- **Recommendation 1.**

**The use of a single expert reviewer in diagnostic ultrasound quality assurance does not provide adequate assurance in terms of inter-rater reliability and therefore should not be used. A quality assurance program based on retrospective assessment of ultrasound imaging and clinical reporting should be undertaken by several reviewers to minimise the effects of inter-reviewer variation.**

### **8.3 Systematic differences between groups.**

There was evidence of weak variability based on systematic differences between study participants. The data shows that those of a greater experience rate image quality more highly when compared with those of lesser experience. Conversely, those of a *lower* clinical grade tend to rate the quality of clinical reports higher than those of a higher clinical grade. No significant differences could be found between those study participants considered 'expert' when compared with those classified as 'non-expert'. These results open the question about which groups of staff are best placed to undertake quality assurance in ultrasound on a day-to-day basis at ultrasound department level.

While some practitioners advocate peer audit as a suitable mechanism of quality assurance[112], there is the finding that those of lesser experience and grading tend to rate both ultrasound imaging and reports differently from their more experienced and more highly graded colleagues. While peer audit may be beneficial at individual practitioner level, unsupervised peer review is unlikely to offer department results which are robust in terms of both quality control and quality improvement. It is therefore recommended that audit for quality assurance programs is either undertaken by, or supervised by those of a higher clinical grade than those whose work is being appraised.

- **Recommendation 2.**

**Quality assurance of non-obstetric ultrasound examinations should ideally be undertaken by those of a higher clinical grade than those undertaking the work being appraised. Peer audit may be acceptable providing there is effective oversight by a senior clinical practitioner.**

#### **8.4 Use of the internet as a mechanism of undertaking quality assurance and disseminating feedback.**

Use of a specially designed website has enabled this project to be successfully completed in that it has facilitated multiple study participants to review, appraise and comment upon many ultrasound images and reports at a time convenient to them. There has been considerable cost saving in that it has not been necessary to copy large amounts of ultrasound imaging and paperwork and send it to study participants by post. Instead, there was only been the requirement to upload the relevant imaging onto a website, enabling easy study participation with minimal fuss. The advantages of using the internet for this work should not therefore be underestimated.

In terms of offering feedback, it must be recognised that there are also significant disadvantages to undertaking this work on-line rather than locally. The internet provides a rather 'sterile' environment in which to offer both quantitative and qualitative feedback. Evidence has already been examined which demonstrates that because of the rather impersonal nature of internet feedback, it is less likely to be

effective than feedback offered face-to-face. The challenge is therefore to utilise fully those aspects of the internet which reduce barriers to performing robust ultrasound quality assurance while minimising the negative aspects of offering feedback and criticism on-line.

It is therefore suggested that the senior practitioner responsible for quality assurance should be responsible for receiving and collating feedback from reviewers external to the department. This practitioner should be responsible for ensuring that this feedback is communicated to individual practitioners and ultrasound units in a way that is constructive and most likely to result in positive clinical outcomes.

- **Recommendation 3.**

**Use of the internet provides great advantages in terms of overcoming logistical difficulties in undertaking quality assurance in ultrasound, particularly when undertaken by external reviewers. However, individual feedback should be given to ultrasound practitioners face-to-face by the senior practitioner responsible for the quality assurance program.**

### **8.5 Qualitative feedback by reviewers.**

This findings of this study are that a significant amount of criticism which was subjective in nature, could be classed as non-constructive when judged against evidence-based and objective criteria[87]. As previously described, the way that criticism is offered is critical in the outcome of that criticism. While the positive

effects of constructive criticism may be small, there is the potential that clumsily given feedback can result in a negative outcome and resultant deterioration in clinical standards. Careful consideration of tone and content of qualitative feedback must therefore be considered an essential aspect of effective feedback mechanisms, particularly in view of its subjective nature.

While analysis of the comments by a specialist in diagnostic ultrasound concluded that expert practitioners gave more constructive feedback than non-experts, this result could have been biased by the content of that feedback in terms of its technical accuracy and veracity. To mitigate this potential bias, a health-care professional, but non-specialist in ultrasound was also asked to analyse the same comments using the same criteria. This analysis contradicted that of the specialist practitioner in that no significant difference was found in the constructiveness of comments given by experts when compared with non-experts in ultrasound. The result implies that the tone of qualitative feedback is of equal importance to the technical content of that feedback.

There was no difference found between expert and non-expert ultrasound practitioners in their ability to offer qualitative comments which were constructive in terms of language and tone of those comments. Most evidence suggests that suitable theoretical and practical training into giving feedback has a significant effect on the quality of that feedback. While it remains unknown as to whether this has a beneficial effect on actual clinical outcomes, it would seem safe to assume that

training in giving feedback and criticism would have a beneficial outcome on the way that feedback is received.

- **Recommendation 4.**

**Those undertaking quality assurance work, regardless of clinical grade and expertise should receive formal training in giving feedback. This is to maximise the potential for this feedback to lead to improved clinical standards and outcomes for patients.**

### **8.6 Selection of Audit Tools.**

Several audit tools were assessed during this project. As expected, there was correlation between all audit tools. There was no significant difference in terms of inter-rater agreement between audit tools. These findings suggest that the specific tool used for quality assurance purposes is of little importance. However, in terms of benchmarking of standards nationally, the use of a nationally recognised audit tool would be of great help. There are several tools which are freely available[26, 112] but there is no primary research evidence to favour selection of one tool over another. In the absence of national guidance, it would be helpful for individual ultrasound units to select an existing freely available audit tool which best suits the workload of their department.

- **Recommendation 5.**

**There is currently no primary research evidence to favour one quality assurance tool over another. Audit tool selection may therefore be done at an individual ultrasound unit level. In selection of a quality assurance tool,**

**ultrasound units should consider which tool best reflects the individual requirements and workload of that unit.**

# **Chapter 9.**

## **Recommendations for Further Research.**

## **Chapter 9.**

### **Recommendations for Further research.**

**This chapter describes areas where study results have uncovered gaps in the knowledge base. It makes recommendations where further research may be warranted.**

#### **9.1 Introduction.**

This study has been a very useful exercise in bringing a degree of objectivity into the principles and practice of clinical quality assurance in diagnostic ultrasound imaging. It has introduced several new concepts, such as use of the internet to disseminate ultrasound imaging to external reviewers and collate results of that imaging.

Because of the pilot nature of this work[36], this project has opened up several avenues that require further exploration and clarification. These are summarised with recommendations for further research activity.

#### **9.2 Use of the internet at a tool to improve clinical outcomes in diagnostic ultrasound.**

This project has made extensive use of the internet in both disseminating and collating reviews on ultrasound imaging from study participants. As previously described, there are powerful logistic and financial advantages gained when utilising this methodology. This study did not address how feedback offered on-line could potentially cause change in behaviour by the recipients of such reviews and how such behavioural change affect clinical outcome measures.

- **Recommendation 1.**

**A detailed evaluation of feedback mechanisms on-line should be undertaken to ascertain the best method of giving on-line feedback in ultrasound quality assurance. This should include objective comparison between different mechanisms of internet feedback and their effect on subsequent recipient behaviour and clinical outcomes.**

### **9.3 Benefits of formal training in giving feedback to reviewers of ultrasound imaging.**

While there is compelling evidence to suggest that receiving formal theoretical and practical training in giving feedback has a positive influence on subsequent feedback cycles, there appears to be little training offered to sonographers in this field. This is despite the considerable difficulties in undertaking robust and objective quality assurance procedures in this specialist field. It is likely (based on existing literature) that formal education will have a positive effect on the constructiveness of feedback given, but this assumption should be tested together with formal measurement of magnitude of the effect.

- **Recommendation 2.**

**Research should be undertaken to examine the effects of formal education in giving feedback to practitioners undertaking ultrasound quality assurance procedures. This research should evaluate the effects of training on subsequent reception of feedback to ultrasound practitioners.**

#### **9.4 Detailed modelling of the effects of participant experience on quality assurance scores given.**

One of the findings of this study was that the length of participant experience influences the way that image quality is assessed and modelled. This study did not collect data of sufficient power to allow detailed analysis of this phenomenon but there is the suggestion that the magnitude of effect may be in the region of 23% when practitioners of no experience and 40 years' experience are compared.

However, this figure has been extrapolated from the data and assumes a uniform, linear change in scores given against length of clinical experience which is probably not present in individual practitioners. It is more likely that the effect is more marked in the initial years of an individual's career and then tails off when maximum confidence and competence is reached.

- **Recommendation 3.**

**The effects of practitioner experience on scores given for image quality should be formally modelled. This should determine whether it is truly linear in nature throughout a practitioner's career or whether there is a plateau, beyond which years of experience have no further effect on scores given for image quality.**

#### **9.5 Who should undertake quality assurance?**

It is currently unclear what grade of staff should undertake ultrasound quality assurance procedures. Both peer audit and audit by those of a higher clinical grade have distinct advantages and disadvantages according to the findings of this study.

There is currently insufficient evidence to determine which method has the best effect on clinical outcomes at a department level.

- **Recommendation 4.**

**Research should be undertaken comparing the effects of ultrasound quality assurance processes being undertaken by peers against those undertaken by practitioners of a higher clinical grade. One of the outcome measures should be the effects on measureable clinical outcomes within an ultrasound department.**

## **Chapter 10.**

### **Study Limitations and Research Learning Opportunities.**

## Chapter 10.

### Study limitations and Research Learning Opportunities.

**This chapter seeks to identify potential flaws within the study design, methodology and results. The potential confounding effect of these flaws and opportunities for learning by the principal researcher are described.**

#### 10.1 Introduction

This chapter seeks to examine some of the limitations of this study. It is intended to examine areas of potential bias and to assess the methodology for potential flaws which may lead to errors in results or interpretation of those results. It also seeks to examine in retrospect some potential weaknesses of the study design and methodology which may have limited this study's scope and potential.

The purpose of such reflection is first to highlight areas of methodological weakness which can be accounted for in the results and interpretation of study data. Such reflection is also an essential aspect in the development of research skills of the principal researcher. While some oversights and errors are inevitable in a study of this scope, honest reflection will help to shape the research skills of those involved to ensure that such flaws can be anticipated in research studies in the future.

The processes and methodology of the quality assurance described in this project clearly have flaws as demonstrated by the lower than expected levels of inter-rater agreement. Alternative methodology such as use of ultrasound simulators were considered as tools to undertake this project. While such methodology may assist in standardisation of processes and methodology, such methods were not considered

applicable in the 'real-world'. A quality assurance program, completely reliant on use of simulators would be unlikely to adjust to the real-world uncertainties and heterogeneity of the work undertaken in most ultrasound departments on a daily basis. The use of simulators (in this particular situation) would focus attention away from overall department performance with regards to real patients, in favour of assessing the performance of individual practitioners as measured by an ultrasound simulator. The methodology described in this project is necessarily imprecise because of the imprecise and subjective nature of the work undertaken. One of the key findings of this study is that such ambiguity is inherent within the imaging modality itself and it is therefore necessary to manage these uncertainties rather than try and eliminate them. While it is accepted that ultrasound simulators have very real benefits in terms of training, their use in measuring a department's diagnostic performance is likely to be very limited.

## **10.2 Bias.**

Bias describes systematic errors in study methodology and interpretation and they 'recur predictably in particular circumstances'[116]. Biases which may be particularly applicable to this study include selection and transition biases[117]

### Selection bias.

Selection bias is characterised by an 'absence of comparability between groups being studied[118]'.

#### 10.2.1 Self-Selection

There is the likelihood of self-selection bias within this study in that only 5 study participants completed all the required reviews and could be included in the inter-rater agreement statistics. It is unlikely that these five study participants are truly representative of the general population of ultrasound practitioners with the UK. Given the time commitment required for this study, it is more likely that these 5 reviewers are different in some way from their counterparts.

It may be that those sonographers completing all reviews were less clinically active, and therefore had more time to take part in research work. Alternatively, this group of ultrasound practitioners may have been more dedicated to their work and hence more inclined to spend the time in undertaking ultrasound study reviews for research purposes. If this were the case, it would be reasonable to assume that the reviews offered by this group were of high quality, particularly given the participants' willingness to take part in research. It could be assumed (but not proven) that inter-rater agreement among this group of ultrasound practitioners would be at least as good as that of the general population of sonographers. More likely, it would be better than that of ultrasound practitioners who may not be as willing to spend as much time undertaking scan reviews for research purposes.

It is unclear to what extent this self-selection bias has affected the results obtained during this study. However, it is assumed that those ultrasound practitioners dedicated enough to spend time on research would produce reviews of high quality and would not have inter-rater variation significantly lower than that of the general population of ultrasound practitioners.

#### 10.2.2 Attrition bias

In total, 17 ultrasound practitioners agreed to take part in this research project on a self-selection basis, yet only 5 (29%) completed all the required reviews of ultrasound imaging. 12 participants completed a varying number of reviews, but because of data analysis limitations, inter-observer agreement could only be judged between these 5 participants. It is unclear in this study data whether there were systematic reasons for attrition in the remaining study participants. The effect of this attrition on the validity of the study results is therefore difficult to assess due to incomplete data.

### 10.2.3 Transfer Bias.

Transfer bias defines how study participants are treated if lost to follow-up, or in this case do not complete the required number of study reviews[119]. Whether individuals not completing all the study reviews are in some way different to the general population of ultrasound practitioners needs to be considered. Review of the data demonstrates that the clear majority of reviews were completed by (non-medically qualified) sonographers. Conversely, medically qualified radiologists, completed only a small number of reviews of the available ultrasound imaging. While the reviews undertaken by radiologists have been included in comparisons of means tests, they have not been included in either the intra-class correlation or logistic regression. The results of this study cannot therefore be applied to radiologists. It is likely that their different professional background cause them to work in a different way from non-medically qualified ultrasound practitioners. They may well therefore have different intra-class correlation scores and regression

results. This finding of this study can therefore only be applied to sonographers and not radiologists.

#### 10.2.4 The Halo Effect

The halo effect describes how judgements (negative or positive) are generalised in their entirety on very limited evidence, rather than making an individual and accurate judgement for each individual element of performance[120]. Because of this cognitive bias, it is possible for a study participant to negatively or positively judge the entire image set based purely on their initial impressions of the first set of ultrasound imaging reviewed. In retrospect, it would have been helpful to have designed the website such that ultrasound examinations were displayed in random order so that the magnitude of the halo effect could have been estimated. Although this bias has not been accounted for within the design phase of this project, it is important that this effect should be borne in mind for subsequent studies.

#### **10.3 Sample size.**

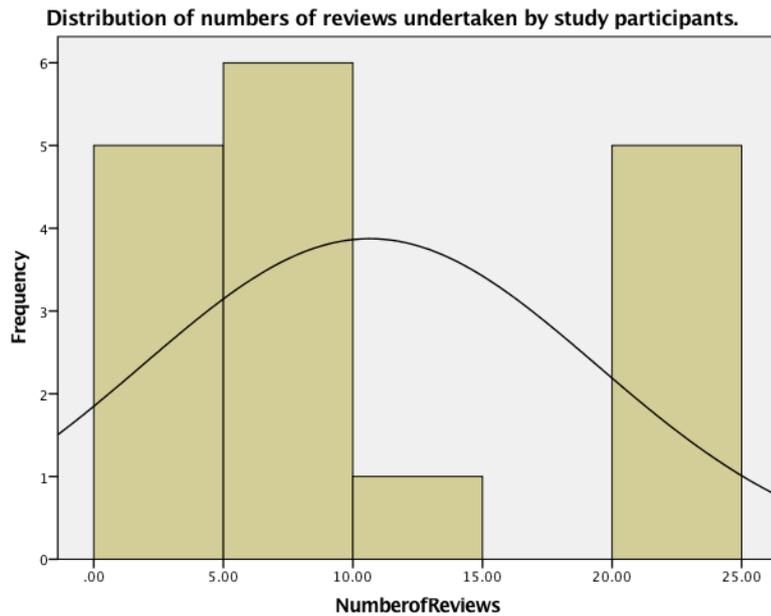
The number of completed reviews was less than anticipated with fewer participants undertaking fewer reviews of ultrasound studies than had been hoped. The initial intention was that the study would have incorporated 50 study participants with each participant reviewing all 22 ultrasound examinations displayed. The number of study participants was significantly less with 17 study participants and only 5 participants who reviewed all the ultrasound examinations displayed. The statistical measures of inter-rater agreement did not permit data from those participants who did not complete all reviews to be included. This could be considered a major flaw within this study and it is therefore necessary to revisit the data to assess whether the smaller

than expected numbers of completed reviews are sufficient to justify the conclusions drawn. The confidence intervals from the intra-class correlation range from 'poor' to 'moderate' and the results do therefore confirm the study findings of much lower than expected levels of agreement between ultrasound practitioners, even based upon this small sample size. While the sample size is small, the confidence intervals do therefore allow a reasonable level of confidence in the overall findings and conclusions of the study. However, it would be prudent to attempt to find additional means of measuring inter-rater agreement which enables the data from those study participants who were unable to complete all reviews to be considered. While the numbers (as presented) are just sufficient for the statistical analysis intended, they do not permit detailed 'drilling-down' into the results and therefore some conclusions cannot be adequately drawn from the results of this study.

This could have been anticipated within the design phase of the project.

Recruitment of subjects into research projects is known to be a challenging area of research which requires careful assessment to maximise the chances of a study reaching its objectives[121-124]. This has been the case for this study also.

Examining the numbers of ultrasound examinations rated by each of the study participants show that rather than all the study participants reviewing all the ultrasound examinations as anticipated, the number of studies reviewed per participant follow an approximately normal distribution as given in Figure 19. This could have been predicted through piloting and accounted for in the design phase of the project in terms of expected numbers of completed ultrasound reviews per study participant.



**Figure 9. Graph demonstrating frequency distribution of number of ultrasound examinations reviewed by study participants**

#### **10.4 Recruitment of study subjects.**

For successful completion of this project, the internet has been an essential tool in ensuring that geographically separate ultrasound practitioners had the opportunity to independently review, score and comment upon the same set of ultrasound examinations. In retrospect, social media could have been utilised in terms of study recruitment also. There already exist social networking groups for ultrasound practitioners which are administered by professional bodies (E.g. British Medical Society) and this study could have been linked to those sites to enhance recruitment and participation recruitment. The use of social media is a well-recognised tool for

recruiting to clinical research projects[125-129]and future research in this area should make use of social media to aid recruitment of study participants.

Use of social media for participant recruitment alone may introduce a degree of selection bias into any further research project. It is currently unclear whether there are any systematic differences between ultrasound practitioners who are users of social media compared with non-users of social media. However, social media has been found to be a potent tool in recruiting participants who are traditionally difficult to reach with other recruitment methods[125, 126]. If used in conjunction with other, more traditional recruitment methods, this would result in a much more representative sample of study participants than that obtained by using the methods in this study alone. There is also the potential for considerable financial saving in recruitment through social media[126].

### **10.5 Intellectual Property.**

The development of a bespoke, web-based platform for review of ultrasound examinations with the ability to collect feedback remotely and in real-time has generated significant intellectual property within this project. Intellectual property describes 'creative work which can be treated as an asset or physical property'[130]. Intellectual property rights fall into four main areas; copyright, trademarks, design rights and patents. There were two aspects of intellectual property which were relevant to this project[131].

### 10.5.1 Copyright

Copyright is automatically assigned to the content of the website and therefore is under the control of the principal researcher. However, it should be remembered that the content of the website was drawn (with permission) from the principal researcher's employing institution and therefore would belong to that institution.

### 10.5.2 Design Rights.

Design rights describe the appearance of a product [132], in this case, the overall design, functionality and look of the website. Because no plan for ownership of design rights was made during collaboration between the principal researcher and website designer, the issue of design rights has proven to be particularly intractable. The website was created by an undergraduate computer scientist as part of a final year university project. No financial payment was made for creation of the website because both the principal researcher and undergraduate computer scientist stood to gain from design of the site within their respective academic programs of study. The website was hosted by the University of Exeter and the website design took place while the principal researcher was employed within an NHS institution. There are therefore multiple legitimate claims on design rights and the issues of intellectual property have become particularly complex.

Advice was sought retrospectively from University and NHS intellectual property experts. Because assignment of intellectual property had not been agreed in advance, the assignment of intellectual property rights has been extremely difficult

and time-consuming. A pragmatic decision was taken to continue with the research study leaving aside issues of intellectual property, in order that the study could be completed with timely dissemination of results.

This has had some negative consequences in that there has been considerable interest in the design of the website and its use in undertaking quality assurance in diagnostic ultrasound. At least one professional body (BMUS) has requested access to, and use of the site for research into professional standards. Because of the unresolved issues of intellectual property, allowing such access has not been possible.

In retrospect, it would have been much better to have resolved the issue of intellectual property before design of the website took place. Application for a research grant could have been made to support the funding of website development, but this would also have required careful management in terms of agreeing intellectual property between the funding institution, university, principal researcher's employing institution and principal researcher himself. Intellectual property expertise is however freely available through both the university and NHS in terms of negotiating and agreeing intellectual property rights and it would have been beneficial to have agreed this at the start of, rather than during the study.

# **Chapter 11**

## **Conclusions and Summary.**

## **Chapter 11**

### **Conclusions and Summary.**

**This short chapter aims to revisit the original objectives of the study and assess how well these have been met. Key findings from the study are presented and a short summary included.**

#### **11.1 Introduction**

This study has been wide-ranging in scope. It has incorporated design and use of bespoke computer software, introduced a novel means of undertaking quality assurance in medical ultrasound and has been the first study to formally undertake testing of existing audit tools and generate knowledge on inter-rater agreement and systematic effects in the use of these tools. It is important to re-examine the original study aims and objectives of this study and assess how well these objectives have been met.

#### **11.2 Original Study Objectives.**

##### **11.2.1 To create, develop and evaluate a standard image bank and audit tool to facilitate standardisation in quality assurance.**

An image bank has been successfully created for the purposes of this project. The principal researcher already possessed considerable data around quality assurance in diagnostic ultrasound within his own clinical department. This enabled an appropriate ultrasound image bank to be created which reflected the whole range of quality of ultrasound imaging to be included across the entire quality spectrum. It

has ensured that data from this study is applicable to ultrasound studies of both high and low clinical quality. Although there is no primary research evidence, anecdotal evidence suggests that the complexity of work for some ultrasound providers (particularly community providers) is low with consequent low yield of pathology. To use a random selection of ultrasound studies from such sources would introduce a serious risk of bias where these 'easier' ultrasound studies were more heavily represented within the image bank than more challenging ultrasound scans. The use of previous internal quality assurance data has ensured that more challenging ultrasound scans could be included. This has ensured that the study is representative of a wider range of ultrasound examination difficulty and quality.

The standardisation of the image set for evaluation by several reviewers is a novel concept in medical ultrasound. It has enabled direct comparison between independent reviewers in the way that they assess the clinical quality of ultrasound examinations. Systematic differences between reviewers can therefore be appreciated using this methodology.

#### **11.2.2 To determine the degree of inter-rater variation in assessment of ultrasound images and reports.**

This study has been successful in determining the level of inter-rater agreement between ultrasound practitioners. The level of agreement was assessed as 'fair' by objective methodology. There is no primary research evidence in this field to benchmark these results against, but peer review of the scientific literature provides

an analogous situation. The results of this study are similar to inter-rater agreement results during the peer review process. This level of inter-rater agreement is too low to recommend that reviews are undertaken by a single ultrasound practitioner and a range of practitioners should therefore be utilised to buffer against inter-rater variation.

This study does have some limitations in assessment of inter-rater agreement. Recruitment and attrition difficulties have hampered the assessment of agreement between reviewers with fewer study participants completing the required number of reviews than anticipated. However, the confidence intervals within the intra-class correlation are sufficiently narrow to have reasonable faith in the validity of this finding.

Assessment of agreement between reviewers was only undertaken among non-medically qualified ultrasound practitioners, so these results cannot be directly applied to medics working in ultrasound (usually radiologists). However, there are no reasons to indicate that inter-rater agreement is different for radiologists than it is for sonographers. However, this has not been formally tested within this study.

### **11.2.3 To determine whether there are systematic variations in data based on reviewer characteristics.**

Diagnostic ultrasound is undertaken by a wide variety of clinical staff and it is therefore important to understand the systematic effects of individual reviewer characteristics so that these can be corrected for within a quality assurance program.

The characteristics investigated in this study were:

- Professional background
- Participant Expertise
- Clinical grade
- Length of experience,
- Age of equipment within the practitioner's employing institution
- Type of review setting in which the ultrasound practitioner practiced.

This study demonstrated some weak systematic effects, particularly in the way that those of greater clinical experience rated image quality and the effects of clinical grade on rating of clinical report quality. Precise modelling of these effects has been complex, particularly in regards to the effect of clinical grade and this has been further hampered by the paucity of data. The prediction of the precise effects therefore requires further evaluation and investigation.

Participant expertise and age of ultrasound equipment used by study participants had no systematic effects on the scores given for study quality. There was insufficient data to analyse systematic effects of professional background and the type of clinical setting in which the study participant practiced.

Overall, the systematic effects of the characteristics of the study participants are relatively weak. A larger cohort of study participants would be required to model these (small) effects more precisely.

#### **11.2.4 To compare different audit tools currently used in quality assurance of non-obstetric diagnostic ultrasound.**

This study was successful in comparing a range of audit tools with inter-rater agreement being the main outcome measure. The tools enabled collection of both ordinal and continuous data. Although the wording and scoring system for each of the tools were different, there was little difference in the levels of inter-rater agreement between the tools. As expected, there was a good correlation between audit tools, implying that the audit tools were all measuring the same thing; the quality of an ultrasound examination.

The results strongly imply that the level of inter-rater variation is inherent in the audit methodology itself, rather than the audit tools used. Acceptance of a degree of inter-rater variability and of the fact that ultrasound quality assurance is not an exact science is therefore important. Management of this variability is likely to be of more value than attempting to eradicate variation between reviewers.

#### **11.2.5 To pilot novel, dynamic IT solutions facilitating distribution and review of such images and enable completion of quality assurance reviews. To explore mechanisms by which quality assurance can be undertaken between departments.**

The development of a suitable platform for image sharing and data collection between geographically separate practitioners has been successfully undertaken. Although issues with feedback given on-line, rather than face-to-face remain, the practical advantages of such a system outweigh the disadvantages. A thorough

grasp of some of the theoretical underpinnings of giving feedback together with formal training for those involved in giving feedback should enable this to be successfully managed. This is particularly important where the feedback is qualitative and subjective.

A major stumbling block within this project has been that of intellectual property. While a suitable platform has been developed and successfully utilised for the purposes of this project, this platform is unlikely to make it into real, clinical use because of intellectual property issues. It is recommended that any further sites are developed only after the issue of intellectual property has been unambiguously settled.

Overall, this study has fulfilled its stated aims. As a completely novel study, this fits neatly into the piloting, feasibility and development stages of research in the MRC complex interventions guidance[36]. There were some flaws in study design and significant issues around participant recruitment. Suggestions on how these factors may be mitigated in any subsequent studies have been given.

The study has enabled a number of evidence-based clinical recommendations to be drawn up with regard to processes in quality assurance of medical diagnostic ultrasound. In addition, it has highlighted areas where further research would be helpful. In this context, this study, while containing some flaws, has been a success.

## **Acknowledgements**

Thanks to the numerous study participants who freely gave their time to undertake the required ultrasound reviews. Thanks also to Justin Turner for his expertise in website design and Dr Obi Ukoumunne for his expertise in statistics.

Thanks to Professor Karen Knapp, my academic supervisor for her frequent guidance, patience and common sense and to my field collaborators, Drs Simon Freeman and Cath Gutteridge for putting up with my constant pestering!

Finally, thanks to my family, Fran, Rosie, Anna and James for giving me the time and space (and frequent cups of coffee) needed to undertake this project.

# References

## References

1. The College of Radiographers, *Medical Image Interpretation and Clinical Reporting by Non-Radiologists: The Role of the Radiographer.* , in *College of Radiographers. London.* . 2005
2. McKenzie G, M.S., Graham D et al, *Radiographer performed general diagnostic ultrasound: current UK practice.* *Radiography*, 2000. **6**: p. 178-188.
3. The Royal College of Radiologists and College of Radiographers, *Team working within clinical imaging. A contemporary view of skills mix.* , in *COR and RCR.* 2007: London.
4. General Medical Council, *Good Medical Practice*, GMC, Editor. 2009: London.
5. The Society and College of Radiographers, *Preliminary Clinical Evaluation and Clinical Reporting by Radiographers: Policy and Practice Guidance.* 2013, SCOR.: London.
6. Department of Health, *Equity and excellence: Liberating the NHS*, D.o. Health, Editor. 2010, HMSO: London.
7. United Kingdom Accreditation Service. 2010; Available from: <http://www.isas-uk.org/>.
8. Schoppler G, H.J., Michaely H et al. , *The impact of ultrasound in urology.* . *Urologe*, 2012. **51**(1): p. 81-98.
9. Tossetti C, B.S., Benedetto E et al, *The management of patients with new-onset of upper gastro-intestinal symptoms in primary care.* *Digestive and Liver Disease* 2010. **42**(12): p. 860-864.
10. R., A.A.a.E., *Practical Head and Neck Ultrasound.* 2000, London: Greenwich Medical Media Ltd.
11. Robinson P, C.G., Wiggins M, *Interpretation of selected accident and emergency radiographic examinations by radiographers: a review of 11000 cases.* *The British Journal of Radiology*, 1999. **72**: p. 546-551.
12. Wivell G, D.E., Eve C et al, *Can Radiographers read screening mammograms?* *Clinical Radiology* 2003. **58**: p. 63-67.
13. Brealey S, S.A., Hahn S et al, *Accuracy of radiographer plain film reporting in clinical practice: a meta-analysis.* *Clinical Radiology* 2005. **60**: p. 232-241.

14. Leslie A, V.J., *Detection of colorectal carcinoma on double contrast barium enema when double reporting is routinely performed: An audit of current practice.* . Clinical Radiology 2002. **57**: p. 184-187.
15. Bossuyt M, R.J., Bruns D et al. , *Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative.* Am J Roentgenol 2003. **181**(1): p. 51-55.
16. Bates, J.A., R.M. Conlon, and H.C. Irving, *An audit of the role of the sonographer in non-obstetric ultrasound.* Clin Radiol, 1994. **49**(9): p. 617-20.
17. Reynolds P, D.R., Cowan F, *Neonatal cranial ultrasound interpretation: a clinical audit.* . Arch Dis Child Fetal Neonatal Ed 2001. **84**: p. F92-F95.
18. Weston MJ, M.A., Slack N, *An audit of radiographer based ultrasound service.* Br J Radiology, 1994. **67**: p. 665-667.
19. Riley, *Musculoskeletal ultrasound: audit of sonographer reporting.* Ultrasound, 2010. **18**(1): p. 36-40.
20. Leslie A, L.H., Virjee J, *Who should be performing routine abdominal ultrasound? A prospective double blind study comparing the accuracy of radiologist and radiographer.* Clinical Radiology 2000. **55**: p. 606-609.
21. Dongola N, G.R., Ward S et al, *Can sonographers offer an accurate upper abdominal ultrasound service in a district general hospital? .* Radiography 2003. **9**: p. 29-33.
22. Bude R, F.J., Lechtanski T, *The use of additionally trained sonographers as ultrasound practitioners. Our first year experience.* . Journal of Ultrasound in Medicine. , 2006. **25**(3): p. 321-327.
23. Riley S, G.C., Chandramohan, *Musculoskeletal ultrasound: audit of sonographer reporting.* Ultrasound, 2010. **18**: p. 36-40.
24. Bude, R.O., J.P. Fatchett, and T.A. Lechtanski, *The use of additionally trained sonographers as ultrasound practitioners: our first-year experience.* J Ultrasound Med, 2006. **25**(3): p. 321-7; quiz 328-30.
25. Kahn C, L.C., Burnside E et al. , *Toward Best Practices in Radiology Reporting.* Radiology 2009. **252**: p. 852-856.
26. The Royal College of Radiologists and The College of Radiographers, *Standards for the provision of an ultrasound service.* . 2014, Royal College of Radiologists.: London.

27. Plumb A, G.F., Khan S. , *Survey of hospital clinicians' preferences regarding the format of radiology reports*. Clinical Radiology 2009. **64**(4): p. 286-394.
28. Schwartz L, P.D., Berk A et al, *Improving communication of diagnostic radiology findings through structured reporting*. Radiology, 2011. **260**: p. 174-181.
29. Grieve F, P.A., Khan S, *Radiology reporting: a general practitioner's perspective*. Br J Radiology 2010. **83**: p. 17-22.
30. Garcea G, M.A., Ong S et al. , *Caveat reporting in ultrasound interpretation of surgical pathology*. J Eval Clin Pract 2010. **16**(1): p. 97-99.
31. S., B., *Measuring the effects of image interpretation: an evaluative framework*. Clinical Radiology 2001. **56**(5): p. 341-347.
32. McCreadie G, O.T., *Eight CT lessons that we learned the hard way: an analysis of current patterns of radiological error and discrepancy with particular emphasis on CT*. Clinical Radiology, 2009. **64**(5): p. 491-9.
33. Reynolds, P.R., R.C. Dale, and F.M. Cowan, *Neonatal cranial ultrasound interpretation: a clinical audit*. Arch Dis Child Fetal Neonatal Ed, 2001. **84**(2): p. F92-5.
34. Scott H, G.A., *Breast screening: PERFORMS identifies key mammographic training needs*. BJR, 2006. **79**(2): p. S127-S133.
35. Dong L, C.Y., Gale A. , *Early identification of substandard breast screening performers*. Breast Cancer Research 2011. **13**(1).
36. Medical Research Council. *Developing and evaluating complex interventions: new guidance*. 2006; Available from: <https://www.mrc.ac.uk/documents/pdf/complex-interventions-guidance/>.
37. The Kings Fund. 2012; Available from: [http://www.kingsfund.org.uk/current\\_projects/general\\_election\\_2010/frequently\\_asked.html-ahow\\_many\\_hospitals\\_are\\_there\\_in\\_england](http://www.kingsfund.org.uk/current_projects/general_election_2010/frequently_asked.html-ahow_many_hospitals_are_there_in_england).
38. National Health Service. *Diagnostic Tests Direct Access Non-Obstetric Ultrasound Service*. 2013; Available from: [http://www.sor.org/system/files/article/201407/refreshed\\_implementation\\_nous\\_2013](http://www.sor.org/system/files/article/201407/refreshed_implementation_nous_2013).
39. Cantin, P. and K. Knapp, *Diagnostic image quality in gynaecological ultrasound: Who should measure it, what should we measure and how?* Ultrasound, 2013: p. 1742271X13511242.

40. Prochaska J, R.R., Schroeder S et al. , *An online survey of tobacco use, intentions to quit, and cessation strategies among people living with bipolar disorder*. *Bipolar Disorders*, 2011. **13**(5): p. 466-473.
41. MsSherry W, J.S., *An online survey of nurses' perceptions of spirituality and spiritual care*. *Journal of Clinical Nursing*, 2011. **20**(11): p. 1757-1767.
42. Burls A, E.J., ,Quinn T et al. , *Oxygen use in acute myocardial infarction: an online survey of health professionals' practice and beliefs*. *Emerg Med J*, 2010. **27**(283-286).
43. Pope, J., *Medical Physics: Imaging*. 1999: Heinemann.
44. The College of Radiographers, *Ultrasound examination times and appointments*. 2015, COR: London.
45. Wu, T.-C., et al., *An Economical, Personal Computer-based Picture Archiving and Communication System*. *RadioGraphics*, 1999. **19**(2): p. 523-530.
46. Rohatgi, S., et al., *After-Hours Radiology: Challenges and Strategies for the Radiologist*. *American Journal of Roentgenology*, 2015. **205**(5): p. 956-961.
47. The College of Radiographers, *Developing and Growing the Sonographer workforce: Education and Training needs*. 2009, COR: London.
48. British Medical Ultrasound Society, *Proceedings of the British Medical Ultrasound Society 46th Annual Scientific Meeting 9–11 December 2014, The Point, Emirates Old Trafford, Manchester, UK*. *Ultrasound*, 2015. **23**(2).
49. Orlando Catalano, A.N., Alfredo Siani, *Fundamentals in Oncologic Ultrasound: Sonographic Imaging and Intervention in the Cancer Patient*. 2009, Milan: Springer.
50. Gisev, N., J.S. Bell, and T.F. Chen, *Interrater agreement and interrater reliability: Key concepts, approaches, and applications*. *Research in Social and Administrative Pharmacy*. **9**(3): p. 330-338.
51. Hayes, A.F., *Statistical mediation analysis with a multicategorical independent variable*. *British Journal of Mathematical and Statistical Psychology*, 2014. **67**(3): p. 451-470.
52. Podichetty, V., et al., *Assessment of internet use and effects among healthcare professionals: a cross sectional survey*. *Postgraduate Medical Journal*, 2006. **82**(966): p. 274-279.

53. Treweek, S., et al. *Strategies to improve recruitment to randomised controlled trials*. Cochrane Database of Systematic Reviews, 2010. DOI: 10.1002/14651858.MR000013.pub5.
54. Bonevski, B.M., Parker; Horton, Graeme; Foster, Mark and Girgis, Afaf, *Response Rates in GP Surveys: Trialling Two Recruitment Strategies* Australian Family Physician, 2011. **40**(6): p. 427-430.
55. Eysenbach, G. and J. Wyatt, *Using the Internet for Surveys and Health Research*. J Med Internet Res, 2002. **4**(2).
56. Julia V Cook, H.O.D., Martin P Eccles , *Response rates in postal surveys of healthcare professionals between 1996 and 2005: An observational study*. BMC Health Services Research, 2009. **9**: p. 160.
57. Baethge, C., J. Franklin, and S. Mertens, *Substantial Agreement of Referee Recommendations at a General Medical Journal – A Peer Review Evaluation at Deutsches Ärzteblatt International*. PLoS ONE, 2013. **8**(5): p. e61401.
58. Kirsty Kiezebrink, I.K.C., Linda Irvine, Vivien Swanson, Kevin Power, Wendy L Wrieden and Peter W Slane, *Strategies for achieving a high response rate in a home interview survey*. BMC Medical Research Methodology, 2009. **9**: p. 46.
59. McAlindon, T., et al., *Conducting clinical trials over the internet: feasibility study*. BMJ, 2003. **327**(7413): p. 484-487.
60. Medical Research Council. *Personal Information in Medical Research*. 2003; Available from: <http://www.mrc.ac.uk/documents/pdf/personal-information-in-medical-research/>.
61. Orton-Johnson, K., *Ethics in Online Research; Evaluating the ESRC Framework for Research Ethics Categorisation of Risk*. Sociological Research Online, 2010. **15**(4): p. 13.
62. Kim, H.-Y., *Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis*. Restorative Dentistry & Endodontics, 2013. **38**(1): p. 52-54.
63. Altman, D.G., *Practical Statistics for Medical Research*. 1997: Chapman and Hall
64. The Royal College of Radiologists and The Society and College of Radiographers, *Quality assurance in radiology reporting: peer feedback*. 2014, RCR: London.

65. Opinion., *Bad Peer Reviewers*. Nature, 2201. **413**(93).
66. Lindsey, D., *Assessing precision in the manuscript review process: A little better than a dice roll*. Scientometrics, 1988. **14**(1): p. 75-82.
67. Lush, B., *Peer review: Unreliable reviewers*. Journal of the Royal Society of Medicine, 2006. **99**(8): p. 385.
68. Black N, v.R.S., Godlee F, Smith R, Evans S., *What Makes a Good Reviewer and a Good Review for a General Medical Journal?*. JAMA, 1998. **280**(3): p. 231-233.
69. Callaham, M.L. and J. Tercier, *The Relationship of Previous Training and Experience of Journal Peer Reviewers to Subsequent Review Quality*. PLoS Med, 2007. **4**(1): p. e40.
70. Baxt, W.G., et al., *Who Reviews the Reviewers? Feasibility of Using a Fictitious Manuscript to Evaluate Peer Reviewer Performance*. Annals of Emergency Medicine. **32**(3): p. 310-317.
71. Bornmann, L., R. Mutz, and H.-D. Daniel, *A Reliability-Generalization Study of Journal Peer Reviews: A Multilevel Meta-Analysis of Inter-Rater Reliability and Its Determinants*. PLoS ONE, 2010. **5**(12): p. e14331.
72. Publishing, S. *Purpose of Peer Review*. Accessed September 2016; Available from: <https://uk.sagepub.com/en-gb/eur/purpose-of-peer-review>.
73. Peninsula Ultrasound. *Information for General Practitioners*. Accessed September 2016 [cited 2016; Available from: <http://peninsulaultrasound.org/nhs-gps/>].
74. Smith, R., *Peer review: a flawed process at the heart of science and journals*. Journal of the Royal Society of Medicine, 2006. **99**: p. 178-182.
75. Schroter Sara, B.N., Evans Stephen, Carpenter James, Godlee Fiona, Smith Richard et al., *Effects of training on quality of peer review: randomised controlled trial*. BMJ, 2004. **328**: p. 673.
76. Houry, D., S. Green, and M. Callaham, *Does mentoring new peer reviewers improve review quality? A randomized trial*. BMC Medical Education, 2012. **12**: p. 83-83.
77. Kliever, M.A., et al., *Peer Review at the American Journal of Roentgenology: How Reviewer and Manuscript Characteristics Affected Editorial Decisions on 196 Major Papers*. American Journal of Roentgenology, 2004. **183**(6): p. 1545-1550.

78. Mahoney, M.J., *Publication prejudices: An experimental study of confirmatory bias in the peer review system*. Cognitive Therapy and Research, 1977. **1**(2): p. 161-175.
79. Alam, M., et al., *Blinded vs. unblinded peer review of manuscripts submitted to a dermatology journal: a randomized multi-rater study*. British Journal of Dermatology, 2011. **165**(3): p. 563-567.
80. van Rooyen, S., et al., *Effect of open peer review on quality of reviews and on reviewers' recommendations: a randomised trial*. BMJ, 1999. **318**(7175): p. 23-27.
81. Robert A. McNutt, A.T.E., Robert H. Fletcher, Suzanne W. Fletcher, , *The Effects of Blinding on the Quality of Peer Review A Randomized Trial*. JAMA, 1990. **263**(10): p. 1371-1376.
82. Baggs, J., Broome M, Dougherty M ,Freda M, Kearney M., *Blinding in peer review: the preferences of reviewers for nursing journals*. Journal of Advanced Nursing , 2008. **64**(2): p. 131-138.
83. Marsh H, B.N., Jayasinghe, *Peer review process: Assessments by applicant-nominated referees are biased, inflated, unreliable and invalid*. Australian Psychologist, 2007. **42**(1): p. 33-38.
84. Fogelholm, M., et al., *Panel discussion does not improve reliability of peer review for medical research grant proposals*. Journal of Clinical Epidemiology, 2012. **65**(1): p. 47-52.
85. Li, L., X. Liu, and A.L. Steckelberg, *Assessor or assessee: How student learning improves by giving and receiving peer feedback*. British Journal of Educational Technology, 2010. **41**(3): p. 525-536.
86. McLeod, R.S., *Quality Improvement Initiatives in Colorectal Surgery: Value of Physician Feedback*. Diseases of the Colon & Rectum, 2017. **60**(2): p. 133-134.
87. Hamid Y, M.S., *Understanding constructive feedback: A commitment between teachers and students for academic and professional development*. J Pak Med Assoc, 2010. **60**(3): p. 224-227.
88. Howard, C.D., A.F. Barrett, and T.W. Frick, *Anonymity to Promote Peer Feedback: Pre-Service Teachers' Comments in Asynchronous Computer-Mediated Communication*. Journal of Educational Computing Research, 2010. **43**(1): p. 89-112.

89. Bingham, C.M., et al., *The Medical Journal of Australia internet peer-review study*. *The Lancet*, 1998. **352**(9126): p. 441-445.
90. Lee, C.J., et al., *Bias in peer review*. *Journal of the American Society for Information Science and Technology*, 2013. **64**(1): p. 2-17.
91. Hysong, S.J., *Meta-Analysis: Audit & Feedback Features Impact Effectiveness on Care Quality*. *Medical care*, 2009. **47**(3): p. 356-363.
92. Kluger A, D.A., *The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory*. *Psychological Bulletin*, 1996. **119**(2): p. 254-284.
93. Vallacher R, W.D., *Levels of Personal Agency: Individual Variation in Action Identification*. *Journal of Personality and Social Psychology*, 1989. **57**(4): p. 660-671.
94. Ende, J., *Feedback in Clinical Medical Education*. . *JAMA*, 1983. **250**(6): p. 777-781.
95. Branch, W.T.J. and A. Paranjape, *Feedback and Reflection: Teaching Methods for Clinical Settings*. *Academic Medicine*, 2002. **77**(12, Part 1): p. 1185-1188.
96. Ivers, N., et al., *Audit and feedback: effects on professional practice and healthcare outcomes*. *Cochrane Database of Systematic Reviews*, 2012(6).
97. Fowler, P. and B. Wilford, *Formative feedback in the clinical practice setting: What are the perceptions of student radiographers?* *Radiography*, 2016. **22**(1): p. e16-e24.
98. Gude, W.T., et al., *Effect of a web-based audit and feedback intervention with outreach visits on the clinical performance of multidisciplinary teams: a cluster-randomized trial in cardiac rehabilitation*. *Implementation Science : IS*, 2016. **11**: p. 160.
99. van Engen-Verheul, M.M., et al., *Evaluating the effect of a web-based quality improvement system with feedback and outreach visits on guideline concordance in the field of cardiac rehabilitation: rationale and study protocol*. *Implementation Science : IS*, 2014. **9**: p. 780.
100. Lu, J. and N. Law, *Online peer assessment: effects of cognitive and affective feedback*. *Instructional Science*, 2012. **40**(2): p. 257-275.

101. Achterkamp, R., H.J. Hermens, and M.M.R. Vollenbroek-Hutten, *The influence of success experience on self-efficacy when providing feedback through technology*. Computers in Human Behavior, 2015. **52**: p. 419-423.
102. Redwood, S., et al., *Effects of a computerized feedback intervention on safety performance by junior doctors: results from a randomized mixed method study*. BMC Medical Informatics and Decision Making, 2013. **13**(1): p. 63.
103. Al Wahbi, A., *The need for faculty training programs in effective feedback provision*. Advances in Medical Education and Practice, 2014. **5**: p. 263-268.
104. Junod Perron, N., et al., *Effectiveness of a training program in supervisors' ability to provide feedback on residents' communication skills*. Advances in Health Sciences Education, 2013. **18**(5): p. 901-915.
105. Telio, S., R. Ajjawi, and G. Regehr, *The "Educational Alliance" as a Framework for Reconceptualizing Feedback in Medical Education*. Academic Medicine, 2015. **90**(5): p. 609-614.
106. Thrasher, A.B., et al., *Supervising Athletic Trainers' Perceptions of Graduate Assistant Athletic Trainers' Professional Preparation*. Athletic Training Education Journal, 2015. **10**(4): p. 275-286.
107. Van den Bergh, L., A. Ros, and D. Beijaard, *Improving Teacher Feedback During Active Learning*. American Educational Research Journal, 2014. **51**(4): p. 772-809.
108. Harrison, G., *Ultrasound clinical progress monitoring: Who, where and how?* Ultrasound, 2015. **23**(4): p. 197-203.
109. Vafaei, A., et al., *Role of Feedback during Evaluation in Improving Emergency Medicine Residents' Skills; an Experimental Study*. 2016, 2016. **5**(1).
110. Jamtvedt, G., et al., *Audit and feedback: effects on professional practice and health care outcomes*. Cochrane Database of Systematic Reviews, 2006(2).
111. Ibrahim, J., et al., *Interns' perceptions of performance feedback*. Medical Education, 2014. **48**(4): p. 417-429.
112. Parker, P. and O. Byass, *Successful implementation of a performance-related audit tool for sonographers*. Ultrasound, 2015. **23**(2): p. 97-102.
113. Toner, K., M. Gan, and M.R. Leary, *The Impact of Individual and Group Feedback on Environmental Intentions and Self-Beliefs*. Environment and Behavior, 2014. **46**(1): p. 24-45.

114. Ertmer, P.A., et al., *Using Peer Feedback to Enhance the Quality of Student Online Postings: An Exploratory Study*. Journal of Computer-Mediated Communication, 2007. **12**(2): p. 412-433.
115. Aasland, O.G. and R. Førde, *Impact of feeling responsible for adverse events on doctors' personal and professional lives: the importance of being open to criticism from colleagues*. Quality and Safety in Health Care, 2005. **14**(1): p. 13-17.
116. D, K., *Thinking Fast and Slow*. 2011, London: Penguin.
117. Califf, R.M., J. McCall, and R.A. Harrington, *Assessing research results in the medical literature: Trust but verify*. JAMA Internal Medicine, 2013. **173**(12): p. 1053-1055.
118. Grimes, D.A. and K.F. Schulz, *Bias and causal associations in observational research*. The Lancet, 2002. **359**(9302): p. 248-252.
119. Pannucci, C.J. and E.G. Wilkins, *Identifying and Avoiding Bias in Research*. Plastic and reconstructive surgery, 2010. **126**(2): p. 619-625.
120. Rosenzweig, P., *The Halo Effect: . . . and the Eight Other Business Delusions That Deceive Managers*. 2007, New York: Free Press.
121. Bower, P., et al., *Improving recruitment to health research in primary care*. Fam Pract, 2009. **26**.
122. Bower, P., et al., *Interventions to improve recruitment and retention in clinical trials: a survey and workshop to assess current practice and future priorities*. Trials, 2014. **15**(1): p. 399.
123. Fletcher, B., et al., *Improving the recruitment activity of clinicians in randomised controlled trials: a systematic review*. BMJ Open, 2012. **2**(1).
124. Ross, S., et al., *Barriers to Participation in Randomised Controlled Trials*. Journal of Clinical Epidemiology. **52**(12): p. 1143-1156.
125. Kapp, J.M., C. Peters, and D.P. Oliver, *Research Recruitment Using Facebook Advertising: Big Potential, Big Challenges*. Journal of Cancer Education, 2013. **28**(1): p. 134-137.
126. O'Connor, A., et al., *Can I get a retweet please? Health research recruitment and the Twittersphere*. Journal of Advanced Nursing, 2014. **70**(3): p. 599-609.
127. Chetan Khatri, S.J.C., James Glasbey, Michael Kelly, Dmitri Nepogodiev, Aneel Bhangu, J. Edward Fitzgerald, *Social Media and Internet Driven Study*

*Recruitment: Evaluating a New Model for Promoting Collaborator Engagement and Participation.* PLoS ONE 2015. **10**(3).

128. Kaba, A. and T. Beran, *Twelve tips to guide effective participant recruitment for interprofessional education research.* Medical Teacher, 2014. **36**(7): p. 578-584.
129. Arends, I., et al., *How to Engage Occupational Physicians in Recruitment of Research Participants: A Mixed-Methods Study of Challenges and Opportunities.* Journal of Occupational Rehabilitation, 2014. **24**(1): p. 68-78.
130. The UK Copyright Service. [cited 2017 03/03/2017]; Available from: [http://www.copyrightservice.co.uk/copyright/intellectual\\_property](http://www.copyrightservice.co.uk/copyright/intellectual_property).
131. GOV.UK. *Intellectual property and your work.* 03/03/2017]; Available from: <https://www.gov.uk/intellectual-property-an-overview>.
132. UK Copyright Service. *Fact sheet P-15: Designs and design rights.* 2004 [cited 03/03/2017; Available from: [http://www.copyrightservice.co.uk/protect/p15\\_design\\_rights](http://www.copyrightservice.co.uk/protect/p15_design_rights).

# Appendices

## Appendix 1.

### STARD checklist for reporting of studies of diagnostic accuracy.

| Section and Topic           | Item # |   | On page # |
|-----------------------------|--------|---|-----------|
| TITLE/ABSTRACT/<br>KEYWORDS | 1      | Identify the article as a study of diagnostic accuracy (recommend MeSH heading 'sensitivity and specificity').  |           |
| INTRODUCTION                | 2      | State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups.  |           |
| METHODS                     |        |   |           |
| <i>Participants</i>         | 3      | The study population: The inclusion and exclusion criteria, setting and locations where data were collected.  |           |
|                             | 4      | Participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?   |           |
|                             | 5      | Participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in item 3 and 4? If not, specify how participants were further selected.  |           |
|                             | 6      | Data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?  |           |
| <i>Test methods</i>         | 7      | The reference standard and its rationale.   |           |
|                             | 8      | Technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for index tests and reference standard.  |           |
|                             | 9      | Definition of and rationale for the units, cut-offs and/or categories of the results of the index tests and the reference standard.   |           |
|                             | 10     | The number, training and expertise of the persons executing and reading the index tests and the reference standard.   |           |
|                             | 11     | Whether or not the readers of the index tests and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers.   |           |
| <i>Statistical methods</i>  | 12     | Methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g. 95% confidence intervals).   |           |
|                             | 13     | Methods for calculating test reproducibility, if done.  |           |
| RESULTS                     |        |   |           |
| <i>Participants</i>         | 14     | When study was performed, including beginning and end dates of recruitment.   |           |
|                             | 15     | Clinical and demographic characteristics of the study population (at least information on age, gender, spectrum of presenting symptoms).  |           |
|                             | 16     | The number of participants satisfying the criteria for inclusion who did or did not undergo the index tests and/or the reference standard; describe why participants failed to undergo either test (a flow diagram is strongly recommended).        |           |
| <i>Test results</i>         | 17     | Time-interval between the index tests and the reference standard, and any treatment administered in between.  |           |
|                             | 18     | Distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition.   |           |
|                             | 19     | A cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard. |           |
|                             | 20     | Any adverse events from performing the index tests or the reference standard.   |           |
| <i>Estimates</i>            | 21     | Estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals).   |           |
|                             | 22     | How indeterminate results, missing data and outliers of the index tests were handled.   |           |
|                             | 23     | Estimates of variability of diagnostic accuracy between subgroups of participants, readers or centers, if done.   |           |
|                             | 24     | Estimates of test reproducibility, if done.   |           |
| DISCUSSION                  | 25     | Discuss the clinical applicability of the study findings.   |           |

**Appendix 2.**

**Data Extraction Sheet.**

**Name of Study and Date**

**Aims.**

**Recruitment Details.**

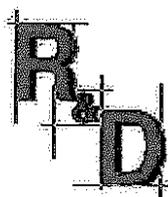
**Type Of Study (Retrospective/Prospective**

**Reference Standard**

**Numbers of participants and Results (with standard deviations)**

**STARD score**

**Appendix 3.**  
**NHS Research and Development Approval.**



**Plymouth Hospitals**   
NHS Trust

Dr Peter Cantin  
Consultant Sonographer  
Derriford Hospital  
Plymouth  
PL6 8DH

Research & Development Office  
Room N17, ITTC Building  
Derriford  
Plymouth  
PL6 8BX

Tel: (01752) 432197/432196  
Fax: (01752) 315 110

05/04/2013

Dear Dr Cantin

**Re: NHS R&D Permission for research project**

**Eudract:** N/A  
**MREC:** N/A  
**UKCRN:** N/A  
**R&D ref:** 13/P/034

Study title: Credible Quality Assurance Procedures in Non-Obstetric Medical Ultrasound. A New Approach to an Old Problem.

This letter confirms that the study named above has Plymouth Hospitals NHS Trust R&D permission to proceed. The governance review carried out included the following documents:

| Document                      | Version    | Document Date |
|-------------------------------|------------|---------------|
| Protocol                      | 1.1        | 16/01/2013    |
| Participant Information Sheet | Appendix 7 | -             |
| Consent Form                  | N/A        | N/A           |

*Note: R&D approval extends to all documents that have received a favourable ethical opinion from the relevant Research Ethics Committee, whether or not they have been referenced in this letter.*

**Please note that from 1<sup>st</sup> April 2011, the Trust's funding is contingent upon research studies recruiting their first patient within 30 calendar days of R&D permission. We therefore encourage you to be in a position to recruit as soon as possible after receipt of this letter.**

Yours sincerely

Dr Lisa Vickers  
R&D Manager

## Appendix 4.

### Caldicott Guardian Approval.

**Plymouth Hospitals**   
NHS Trust

Our ref: ADM/ABB/ PCantin  
03 December 2012

Dr Helen Neilens  
Research Advisor  
Room N17 ITTC Building  
Tamar Science Park  
Derriford  
Plymouth  
PL6 8BX

**South West Cardiothoracic Centre**  
Level 9  
Terence Lewis Building  
Derriford Hospital  
Plymouth  
Devon  
PL6 8DH

**Mr Adrian Marchbank**  
*Consultant Cardiothoracic Surgeon  
and Caldicott Guardian*

Telephone: 01752 431825  
Fax: 01752 763830  
e.mail: [adrian.marchbank@nhs.net](mailto:adrian.marchbank@nhs.net)  
secretary: [angelabradley1@nhs.net](mailto:angelabradley1@nhs.net)

Dear Helen,

**Re: Research Proposal 'Credible Quality Assurance Procedures in Non-Obstetric Ultrasound. A new approach to an old problem.'**

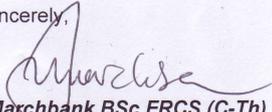
**Researcher: Peter Cantin.**

I have reviewed the above research proposal in my role as Caldicott Guardian. I have also taken advice from Jo Arthur (Information Governance Support Manager) and Jason Scott (ICT Service Development Manager).

This project involves the use of de-identified medical imaging (ultrasound images) to pilot and evaluate new methods of undertaking quality assurance of ultrasound imaging. Procedures for the safe handling and de-identifying of these images are sufficient to protect patient confidentiality. There are sufficient mechanisms in place to ensure that no patient-identifiable data will be available to project participants undertaking review of these images.

I am happy that sufficient regard has been made in protecting the confidentiality of service-user information. I am also satisfied that this research warrants appropriate sharing of de-identified patient imaging.

Yours sincerely,

  
**Mr AJ Marchbank BSc FRCS (C-Th)**  
**Consultant Cardiothoracic Surgeon**

*Working in partnership with the Peninsula Medical School*  
Chairman: Richard Crompton    Chief Executive: Ann James

## **Appendix 5.**

### **Participation Invitation Sheet.**

#### **A study of Quality Assurance in Medical Ultrasound.**

Dear Colleague,

We are conducting a study through Derriford Hospital in Plymouth and the University of Exeter examining ways of improving quality assurance programs in non-obstetric ultrasound.

The study will involve asking ultrasound practitioners of all professional backgrounds and experience to review and grade several ultrasound images and reports using standardised audit tools. Images will be available on-line, but thermal copies or a DVD can be sent to you if you prefer.

It is hoped that this study will provide an insight into variation in ultrasound practices around the country, leading to a better understanding of how we should undertake quality assurance processes within non-obstetric ultrasound.

The audit tools have been designed to allow ultrasound imaging to be reviewed quickly and efficiently. We estimate that reviewing the image set will take approximately 1½ hours. The data will be completely anonymised, but we would be pleased to let you know how your review compares to the overall data if you so wish.

We hope that you will agree to participate in this research.

Yours Sincerely.

Peter Cantin

Principal Investigator.

## Frequently Asked Questions

- **Has this study received ethical approval?**

Yes. This study has been reviewed by both Regional Ethics Committee and University ethics committee and full approval has been granted.

- **Will my image reviews be anonymous?**

All data will be anonymised and coded. Data will be pooled such that no individual results will be analysed unless specifically requested by individual research participants. If you wish to compare your own reviews against that of the research group as a whole, this can be arranged.

- **Do I have to review the images on-line?**

No. The images have been placed on line to facilitate ease of access. However, thermal copies of images or a DVD can be arranged if required.

- **What benefits do I get from the study?**

Quality assurance of non-obstetric ultrasound is an important aspect of commissioning ultrasound services. It is in all providers' interests to develop and use credible, validated methods of undertaking quality assurance. This study will pilot methods of audit, which if successful will be available for all providers to emulate.

- **What level of clinical experience do I need to review images?**

We would like practitioners from all professional backgrounds, and levels of experience to review these images. This will ensure that we know how reviews vary according to these factors.