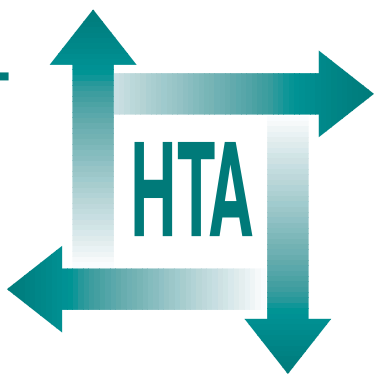
Health Technology Assessment 2001; Vol. 5: No. 27

The cost-effectiveness of magnetic resonance imaging for investigation of the knee joint

S Bryan
G Weatherburn
H Bungay
C Hatrick
C Salas
D Parry
S Field
F Heatley

Health Technology Assessment
NHS R&D HTA Programme



How to obtain copies of this and other HTA Programme reports.
An electronic version of this publication, in Adobe Acrobat format, is available for downloading free of charge for personal use from the HTA website (http://www.hta.ac.uk). A fully searchable CD-ROM is also available (see below).

Printed copies of HTA monographs cost £20 each (post and packing free in the UK) to both public **and** private sector purchasers from our Despatch Agents.

Non-UK purchasers will have to pay a small fee for post and packing. For European countries the cost is £2 per monograph and for the rest of the world £3 per monograph.

You can order HTA monographs from our Despatch Agents:

– fax (with **credit card** or **official purchase order**)
– post (with **credit card** or **official purchase order** or **cheque**)
– phone during office hours (**credit card** only).

Additionally the HTA website allows you **either** to pay securely by credit card **or** to print out your order and then post or fax it.

Contact details are as follows:

HTA Despatch
c/o Direct Mail Works Ltd
4 Oakwood Business Centre
Downley, HAVANT PO9 2NP, UK

Email: orders@hta.ac.uk
Tel: 02392 492 000
Fax: 02392 478 555
Fax from outside the UK: +44 2392 478 555

NHS libraries can subscribe free of charge. Public libraries can subscribe at a very reduced cost of £100 for each volume (normally comprising 30–40 titles). The commercial subscription rate is £300 per volume. Please see our website for details. Subscriptions can only be purchased for the current or forthcoming volume.

Payment methods

Paying by cheque
If you pay by cheque, the cheque must be in **pounds sterling**, made payable to *Direct Mail Works Ltd* and drawn on a bank with a UK address.

Paying by credit card
The following cards are accepted by phone, fax, post or via the website ordering pages: Delta, Eurocard, Mastercard, Solo, Switch and Visa. We advise against sending credit card details in a plain email.

Paying by official purchase order
You can post or fax these, but they must be from public bodies (i.e. NHS or universities) within the UK. We cannot at present accept purchase orders from commercial companies or from outside the UK.

How do I get a copy of *HTA on CD*?

Please use the form on the HTA website (www.hta.ac.uk/htacd.htm). Or contact Direct Mail Works (see contact details above) by email, post, fax or phone. *HTA on CD* is currently free of charge worldwide.

The website also provides information about the HTA Programme and lists the membership of the various committees.

The cost-effectiveness of magnetic resonance imaging for investigation of the knee joint

S Bryan[1,2*]
G Weatherburn[2]
H Bungay[3]
C Hatrick[4]
C Salas[1]
D Parry[1]
S Field[3]
F Heatley[4]

[1] Health Economics Facility, University of Birmingham, UK
[2] Health Economics Research Group, Brunel University, UK
[3] Department of Radiology, Kent and Canterbury Hospital, UK
[4] Orthopaedic Academic Unit, Guy's & St Thomas's NHS Trust, UK

* Corresponding author

Competing interests: none declared

Published July 2001

This report should be referenced as follows:

Bryan S, Weatherburn G, Bungay H, Hatrick C, Salas C, Parry D, *et al*. The cost-effectiveness of magnetic resonance imaging for investigation of the knee joint. *Health Technol Assess* 2001;**5**(27).

Health Technology Assessment is indexed in *Index Medicus*/MEDLINE and *Excerpta Medica*/EMBASE. Copies of the Executive Summaries are available from the NCCHTA website (see opposite).

NHS R&D HTA Programme

The NHS R&D Health Technology Assessment (HTA) Programme was set up in 1993 to ensure that high-quality research information on the costs, effectiveness and broader impact of health technologies is produced in the most efficient way for those who use, manage and provide care in the NHS.

Initially, six HTA panels (pharmaceuticals, acute sector, primary and community care, diagnostics and imaging, population screening, methodology) helped to set the research priorities for the HTA Programme. However, during the past few years there have been a number of changes in and around NHS R&D, such as the establishment of the National Institute for Clinical Excellence (NICE) and the creation of three new research programmes: Service Delivery and Organisation (SDO); New and Emerging Applications of Technology (NEAT); and the Methodology Programme.

This has meant that the HTA panels can now focus more explicitly on health technologies ('health technologies' are broadly defined to include all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care) rather than settings of care. Therefore the panel structure has been redefined and replaced by three new panels: Pharmaceuticals; Therapeutic Procedures (including devices and operations); and Diagnostic Technologies and Screening.

The HTA Programme will continue to commission both primary and secondary research. The HTA Commissioning Board, supported by the National Coordinating Centre for Health Technology Assessment (NCCHTA), will consider and advise the Programme Director on the best research projects to pursue in order to address the research priorities identified by the three HTA panels.

The research reported in this monograph was funded as project number 93/26/16.

The views expressed in this publication are those of the authors and not necessarily those of the HTA Programme or the Department of Health. The editors wish to emphasise that funding and publication of this research by the NHS should not be taken as implicit support for any recommendations made by the authors.

Criteria for inclusion in the HTA monograph series
Reports are published in the HTA monograph series if (1) they have resulted from work commissioned for the HTA Programme, and (2) they are of a sufficiently high scientific quality as assessed by the referees and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search, appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

The editors and publisher have tried to ensure the accuracy of this report but do not accept liability for damages or losses arising from material published in this report. They would like to thank the referees for their constructive comments on the draft document.

ISSN 1366-5278

Published by Core Research, Alton, on behalf of the NCCHTA.
Printed on acid-free paper in the UK by The Basingstoke Press, Basingstoke.

Contents

List of abbreviations

ACL	anterior cruciate ligament
CI	confidence interval
DGH	district general hospital
GP	general practitioner
MR	magnetic resonance
MRI	magnetic resonance imaging
MRS	marginal rate of substitution
RCT	randomised controlled trial
SD	standard deviation
SE	standard error*
VAS	visual analogue scale
SF-36	Short Form with 36 items

* Used only in tables

Executive summary

Objectives

This study considered the role of magnetic resonance imaging (MRI) in the diagnosis of knee injuries in a district general hospital (DGH) setting. The principal objective was to identify whether the use of MRI had a major impact on the clinical management of patients presenting with chronic knee problems, in whom surgery was being considered, whether it reduced overall costs and whether it improved patient outcome.

In addition, the research:

1. explored the 'diagnostic accuracy' of initial clinical investigation of the knee by an orthopaedic trainee, consultant knee specialist and consultant radiologist
2. considered the variability and diagnostic accuracy of interpretations of knee MRI investigations between radiologists
3. measured the strength of preference for the potential diagnostic/therapeutic impact of knee MRI (i.e. the avoidance of surgery).

Methods

Randomised controlled trial

The research was based on a single-centre randomised controlled trial conducted at Kent and Canterbury Hospital. Patients attending with knee problems in whom surgery was being considered were recruited from routine orthopaedic clinics. Most patients had been referred by their general practitioner. Patients were randomised to either investigation using an MRI scan (MRI trial arm) or investigation using arthroscopy (no-MRI trial arm).

The study investigated the benefits of knee MRI at two levels: diagnostic/therapeutic impact (i.e. avoidance of surgery) and patient outcome (using the Short Form with 36 items and EQ-5D quality-of-life measurement instruments). Quality of life was assessed at baseline and at 6 and 12 months. Costs were assessed from the perspectives of the NHS and patients. All analyses were by intention to treat.

Substudies

Investigation of diagnostic accuracy

For the investigation of diagnostic accuracy of initial clinical investigation, the sample comprised 114 patients recruited in a separate study conducted at St Thomas' Hospital. The sample was drawn from patients presenting at the Accident and Emergency Department with an acute knee injury. All study patients received an MRI scan, but initial diagnosis was made without access to the scan or the radiologist's report. After 12 months, all clinical notes and MRI scans of study patients were reviewed and a final 'reference standard' diagnosis for each patient was reached. Comparison was made between the diagnosis recorded by each clinician (i.e. orthopaedic trainee, knee specialist and consultant radiologist) and the reference diagnosis.

Investigation of the generalisability of results

For this substudy, the MRI images from 80 patients (recruited at St Thomas' Hospital) were interpreted independently by seven consultant radiologists at DGHs and the St Thomas' Hospital MRI radiologist. For each area of the knee, the level of agreement (measured using weighted kappa) between the responses of the eight radiologists and the reference standard diagnosis was assessed.

Investigation of preferences

The investigation of potential patient preferences for the diagnostic/therapeutic impact of MRI was explored using a discrete choice conjoint measurement research design. Choices involved selecting between two alternative scenarios described using four attributes, and data were collected from 585 undergraduate sports science students and analysed using a random-effects probit model.

Results

Randomised controlled trial

The trial recruited 118 patients (59 randomly allocated to each arm). The two groups were similar in important respects at baseline.

The central finding was of no statistically significant differences between groups in all

measures of health outcome, although a trend in favour of the no-MRI group was observed. However, the use of MRI was found to be associated with a positive diagnostic/therapeutic impact: a significantly smaller proportion of patients in the MRI group underwent surgery (MRI = 0.41, no-MRI = 0.71; $p = 0.001$). There was a similar mean overall NHS cost for both groups.

Substudies

Investigation of diagnostic accuracy

The exploration of diagnostic accuracy found that, when compared to orthopaedic trainees (44% correct diagnoses) or to radiologists reporting an MRI scan (68% correct diagnoses), the accuracy rate was higher for knee specialists (72% correct diagnoses).

Investigation of the generalisability of results

This generalisability study indicated that, in general terms, radiologists in DGHs provide accurate interpretations of knee MRI images that are similar to a radiologist at a specialist centre. The one area of the knee for which this did not hold was the lateral collateral ligament.

Investigation of preferences

The central finding for this substudy was that, on average and within the range specified, choices in this group of potential patients were not significantly influenced by variation in the chance of avoiding surgery.

Conclusions

Implications for healthcare

The evidence presented in this report supports the conclusions that the use of MRI in patients presenting at DGHs with chronic knee problems in whom arthroscopy was being considered did not increase NHS costs overall, was not associated with significantly worse outcomes and avoided surgery in a significant proportion of patients.

Recommendations for further research (in priority order)

1. The trial data demonstrated that the use of MRI in patients with chronic knee problems reduced the need for surgery. However, the link between diagnostic processes and changes in health outcome is indirect and the finding of no-MRI-related effect on health outcome may, therefore, be a consequence of the limited power of the trial. Further research to confirm (or contradict) these findings would be valuable.
2. The investigation of diagnostic accuracy involved comparison with a reference diagnosis established by a panel of two clinical members of the research team. It would be interesting to explore the extent to which the results would differ using an external panel.
3. The result from the preference study, indicating that the potential diagnostic/therapeutic impact of knee MRI was not highly valued, is a surprising finding that would be important to explore in general public or patient populations.
4. The focus for the trial-based aspects of this research was the DGH and patients presenting with chronic knee problems who were being considered for surgery. Care should be taken in generalising from these results to other patient groups (e.g. acute knee injuries) or to other settings (e.g. specialist centres). Further clinical trials would be required in order to answer such questions.

Chapter 1
Introduction

Since the discovery of X-rays in 1895, which marked the effective birth of diagnostic medical imaging, there have been numerous refinements of imaging techniques and the development of entire new modalities, including ultrasound, computed tomography, magnetic resonance imaging (MRI) and positron emission tomography. As described by Thornbury,[1] the diagnostic imaging process is now a routine part of everyday medical practice whereby imaging technologies are used to record:

> "images of the patient in an imaging medium or system (e.g. [magnetic resonance] MR). These images are then viewed and interpreted by an observer, and diagnostic and prognostic statements are made. The physician managing the patient then takes this information, puts it together with the patient's clinical presentation, laboratory test results and any other relevant information, and makes diagnostic estimates and treatment choices."[1]

Whilst developments in imaging technologies have brought about significant diagnostic improvements and enhancements in the quality of care, the uptake and widespread adoption of these advances has led to diagnostic imaging becoming a major activity in modern medicine that consumes a significant and increasing proportion of the healthcare budget.[2] Economic evaluation can provide guidance on the appropriateness of investments in new (and existing) technologies and, over recent years, there has been a steady growth in the number of published economic analyses in the diagnostic imaging area.[3]

This report focuses specifically on the use of MRI in the diagnosis of knee abnormalities and injuries. The principal research question addressed is:

> In patients presenting at a district general hospital (DGH) with knee problems, and in whom arthroscopy is being considered, does the use of MRI have a major impact on the clinical management of patients, does it bring about an overall reduction in costs and does it improve patient outcome?

Chapter 2 presents the hierarchical framework for the evaluation of diagnostic imaging interventions, originally reported by Fineberg and colleagues[4] in the context of an evaluation of computed tomography, ranging from improvements in diagnostic accuracy through to the effect on health outcome. This framework is then adopted for the research reported here. Chapter 3 provides a brief review of the technology assessment literature concerned with the use of MRI in orthopaedics.

In chapter 4, the results of a randomised controlled trial (RCT) that explored the costs and benefits of using MRI in the diagnosis of knee problems in a DGH are reported. Alongside the trial, an economic evaluation was conducted and this is also reported as part of the same chapter. A number of uncertainties remained following the analysis of the main trial data set, notably:

- the diagnostic accuracy of the initial clinical investigation of knee injuries
- the level of variability and diagnostic accuracy in the interpretation of MRI images by radiologists from non-specialist hospitals (i.e. DGHs) and specialist centres
- the value placed on the diagnostic/therapeutic impacts of knee MRI (i.e. avoidance of surgery) to potential patients.

Each of these issues is then explored empirically in the subsequent chapters (i.e. chapters 5, 6 and 7, respectively).

In the economic evaluation of diagnostic imaging technologies, it is common to employ cost–consequences analyses and the analysis reported in chapter 4 follows this convention. However, this form of analysis has the limitation that in situations where results suggest improvements on some benefit parameters and deterioration on others they do not provide an overall indication of the preferred technology. Whilst cost–utility analysis makes explicit trade-offs between length-of-life and quality-of-life benefits, it does not provide a complete solution for diagnostic imaging technologies, since many of the consequences indicated by Fineberg and colleagues[4] are seen in the short term. Stated preference techniques are able to explore trade-offs for a broader range of benefit parameters.

Therefore, one stated preference approach, known as conjoint measurement, was used here to identify the strength of preferences for the consequences provided by the use of MRI in the context of knee injuries. This study is reported in chapter 7.

Finally, chapter 8 provides a review of all of the research evidence presented in the report and draws conclusions regarding the cost-effectiveness of using MRI in patients presenting with knee problems. In addition, areas requiring further investigation are highlighted.

Chapter 2

The economic evaluation of diagnostic imaging technologies: an overview

Introduction

The focus for this report is the economic evaluation of diagnostic imaging technologies, which constitute a discrete subset of all diagnostic technologies. Whilst there are some distinguishing features of imaging technologies, for example, their tendency to be associated with very high capital costs, the general problems of evaluation discussed here are common to all diagnostic interventions. A central problem encountered by evaluation researchers in this field, as described by Thornbury,[1] is:

> "[the] traditional localised view of the goal of diagnostic radiology...[which is]...that it should provide images of the best technical quality and diagnoses that are as accurate as possible."

Whilst such goals clearly must not be ignored, it is important to maintain an awareness of the position of diagnostic radiology as one part of a larger system that has as its objective the effective and efficient treatment of patients. Adoption of this broader 'systems' view forces one to consider criteria that go beyond technical quality and diagnostic accuracy to an examination of the value to patients and society derived from diagnostic imaging.

The next section briefly discusses the particular difficulties in evaluating diagnostic imaging technologies. The response to these problems by Fineberg and colleagues[4] is then presented and the chapter concludes with a discussion of a possible way forward for economic evaluation in this area.

Health outcome and diagnostic imaging technology evaluation

Many evaluators of diagnostic imaging technologies are sympathetic to the view that a core objective of an economic evaluation of health services, including diagnostic imaging procedures, should be the measurement of effect in terms of health outcome.[1,5] This position is consistent with a belief that the principal rationale for the provision of a publicly funded healthcare system is the improvement of the health of the population. However, as is widely recognised, there are particular difficulties associated with evaluating diagnostic technologies, especially in terms of health outcome.

The central challenge for the evaluation of diagnostic technology concerns the estimation of the technology matrix. This is the relationship between the level of inputs (i.e. the provision of a diagnostic procedure) and the level of outputs (i.e. the improvement in health). This matrix will, in general, be more difficult to estimate for diagnostic interventions compared to therapeutic procedures, since the chain of events by which improvements in the technical capability of diagnostic procedures lead to positive changes in health outcome is likely to be more complex. The use of diagnostic procedures may fail to bring about positive changes in prognosis for some patients through no failing in the diagnostic process but rather through failings of therapy. The link between the diagnostic process and changes in health outcome is indirect; variation in health outcome data will be influenced both by the diagnostic process and by the nature and effectiveness of any treatment provided.[5,6] Thus, in general terms, greater variability will be seen in health outcome data relating to diagnostic technologies compared with such data collected in the evaluation of therapeutic interventions alone. Evaluation studies of diagnostic technologies, therefore, require an appropriate design that takes into account the expected large variability in health outcome data. Such attention to the detail of study design has not been a common characteristic of the evaluation of diagnostic imaging technologies.[7]

Evens[8] emphasised the complex nature of the diagnostic process and the practical implications for evaluation work. He pointed out that many forms of medical diagnoses do not involve a single investigation, but rather combinations of different tests. The implication of this is that the identification of the individual effect of each investigation on health outcome would require

highly complex evaluation designs, with large numbers of test combinations being compared. A further difficulty in attempting to assess the impact of diagnostic technology in terms of health outcome is the narrow definition of health adopted in most conventional measures, with a tendency for the emphasis to be on functional and physical, rather than psychological, aspects of health. The consequence of this is that potentially important aspects of health outcome that might result from the provision of diagnostic information, namely reduced anxiety or increased reassurance, may be inadequately captured using existing measures. Donaldson,[9] Miedzybrodzka and colleagues[10] and Ryan and Shackley[11] have all stressed the need to consider the broader consequences associated with providing diagnostic information. The reassurance derived from negative diagnostic results and, equally, the anxiety created by positive results have tended to be poorly assessed in economic evaluations due to the limited coverage of the measurement instruments available.

Broader scope for benefit assessment: the Fineberg levels

Thus, difficulties exist in exploring the direct relationship between a diagnostic intervention and health outcome. Many of these were recognised by Fineberg and co-workers,[4] who are seen as having significantly moved forward the debate on methods for the evaluation of diagnostic imaging technologies. In the context of evaluating computerised cranial tomography, Fineberg and co-workers[4] asserted that the assessment of effectiveness of diagnostic procedures should distinguish four levels of efficacy (*Figure 1*). Their argument was that evaluations:

> "must do more than define those patients in whom a particular diagnosis is likely to be confirmed or ruled out; it should determine which patients will most likely have therapy altered as a result of the procedure and when a change in therapy will most likely improve outcome. While a change in therapy does not necessarily mean a change in outcome, a diagnostic test can contribute to improved patient outcome primarily because of its effect on the choice and application of therapy."

Whilst Fineberg and colleagues[4] recognised the importance of health outcome, they encouraged evaluators to consider additionally the impact of diagnostic imaging technology on both diagnostic and therapeutic pathways, viewing such change as a possible predictor of change in health outcome. The levels put forward by Fineberg and colleagues[4] have been extended over time and now include 'diagnostic accuracy' as a separate level, which was previously included as part of 'diagnostic impact'. The levels are detailed below and are described with specific reference to imaging technologies.

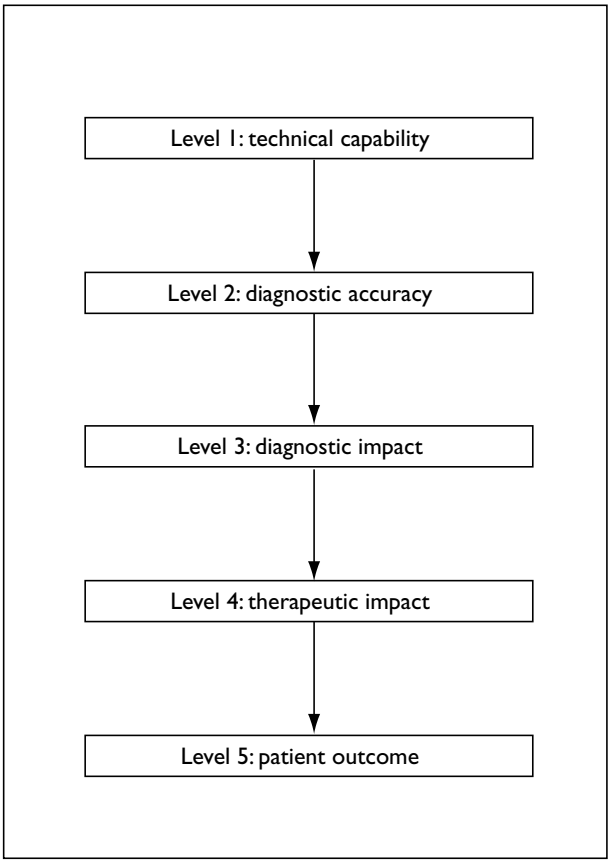


FIGURE 1 *Framework for evaluating diagnostic technologies*

Level 1: technical capability

This is concerned with the basic technical functioning of the technology in question and can be characterised in terms of whether the new imaging equipment produces pictures or images that are technically superior (i.e. physical parameters, such as image resolution and sharpness) to those produced by the alternative diagnostic procedures.

Level 2: diagnostic accuracy

This is sometimes referred to as the 'diagnostic performance' of the technology, and is principally an issue of whether the use of the new imaging equipment brings about an increase in the likelihood of a correct diagnosis being made. Accuracy is commonly reported using two summary statistics: sensitivity (i.e. the proportion of true-positive cases identified by a

test out of all positive cases in a given population) and specificity (i.e. the proportion of true-negative cases identified out of all negative cases). These two measures are sometimes plotted against one another as the threshold for defining a positive case is varied. The plot of sensitivity against 1 – specificity (referred to as a receiver operating characteristic curve) provides a means of making an overall comparison of alternative diagnostic interventions, using the area under the curve as a single indicator of test performance.[1] An assumption that is implicitly made in using this aggregation method is that a single unit change in specificity has the same value as a single unit change in sensitivity; an assumption which seems unlikely to hold in most diagnostic contexts.

Level 3: diagnostic impact

There are two issues of concern here: the extent to which the use of the new technology brings about changes in diagnosis or diagnostic confidence and the consequences associated with the use of the new technology in terms of the path of investigations and tests typically undertaken. The former is based on the premise that if a diagnostic technology does not bring about either a change in patient diagnosis or in the level of confidence associated with diagnosis, then its value should be questioned. The second issue is whether the use of the imaging examination in question causes the nature of subsequent diagnostic processes to change. For example, other diagnostic procedures might be displaced as a result of the use of a new diagnostic technology.

Level 4: therapeutic impact

This level is analogous to diagnostic impact but is concerned with changes in the nature of treatment provided rather than with changes in the diagnostic path. An assessment of therapeutic impact would investigate whether treatment plans are changed once the results of the new diagnostic process are available.

Level 5: patient outcome

The final level is concerned with the impact of the technology on patient outcome. The question that is addressed is: to what extent does the use of the imaging technology in question and related intermediate changes in diagnosis and/or treatment bring about improvements in patient prognosis and final health status?

By placing diagnostic accuracy, diagnostic impact and therapeutic impact between technical capability and patient outcome, Fineberg and colleagues[4] indicate that positive changes in these parameters might be expected to lead to improvements in patient outcome. However, they state that it is not necessary for there to be changes on every efficacy level in order for patient outcome changes to occur. They illustrate their point using diagnostic impact:

> "Changes that computerised cranial tomography may produce in the use of other diagnostic test plans can also directly affect outcome when those other tests carry an intrinsic risk of morbidity or mortality."

As another example, if a new imaging technology was associated with a higher level of radiation dose per examination (i.e. a negative change in technical capability) then the long-term effect on expected patient outcome is likely to be negative, quite possibly without associated intermediate changes in levels 2, 3 or 4. Whilst changes in early levels (such as 2, 3 and/or 4) may bring about an effect in level 5, either directly or indirectly, a change in **any one** of those levels is neither necessary nor sufficient to predict a change in patient outcome. However, a change in patient outcome is unlikely if there is no associated change in any of the other levels and, thus, a change in **at least one** of the earlier levels (levels 1 to 4) is necessary but not sufficient to indicate a change in patient outcome.

Discussion

The levels by Fineberg and co-workers[4] provide a convenient means of classifying the range of consequences that flow from the use of diagnostic imaging technologies. Fineberg and co-workers[4] argued that the consideration of efficacy levels was needed in addition to health outcome since the issues of measurement are more straightforward and changes in levels 1 to 4 are desirable given that they are necessary to indicate change in health outcome. However, changes in levels 1 to 4 are not sufficient to predict health improvements and, without additional information, it is not possible to judge the nature and magnitude of effect on health outcome. Indeed, in some circumstances health outcome may be worse following the use of a new diagnostic technique if, for example, the therapeutic impact was a switch towards more interventionist treatments with associated risks. Thus, focusing exclusively on levels 1 to 4 in an economic evaluation would be inappropriate and possible changes in level 5 cannot be ignored. The implication is that evaluation research of diagnostic imaging technologies needs to consider how patient outcomes can best be incorporated. An obvious first step would be to ensure that

empirical exercises are well designed, taking into account the predictable difficulties in measuring health outcome. For example, in many situations, longer-term patient follow-up may be necessary. Where such problems cannot be addressed through design modifications of the empirical research, modelling of the relationships between the Fineberg levels could usefully be employed.

In addition to patient outcome effects, economic evaluations of diagnostic imaging technologies must also consider changes in the other Fineberg and colleagues' levels because such changes may have important cost and resource-use implications and because they represent consequences that may be highly valued in their own right. An example of the latter is an imaging technology that improves diagnostic accuracy such that confidence in the diagnostic information is raised. The result is the provision of information with greater certainty which may itself confer important benefits on patients, for example, in the form of reduced anxiety levels, even if there is no associated impact on either professional or personal decision-making.[11] In addition, if a diagnostic procedure brings about diagnostic and therapeutic impacts then direct benefits may be experienced by patients if further, possibly invasive, procedures are avoided. Many diagnostic and therapeutic procedures have undesirable consequences for patients, such as pain or additional limitations in performing usual activities over the short-term. In the diagnosis of serious knee problems, for example, an invasive surgical procedure is commonly used that requires hospital admission and general anaesthesia. If it is possible to avoid surgery by using a non-invasive form of diagnosis (such as MRI), all other things remaining equal, then the diagnostic impact would clearly have the effect of conferring benefit on the patient, which, on the assumption of a strong preference to avoid surgery, should be incorporated as a benefit in an economic evaluation. The strength of such preferences is explored in this report.

The implication is that, for economic evaluation purposes, it is necessary to go beyond reporting benefits in a descriptive manner, as in the Fineberg framework, since this does not allow the strength of preference or relative importance of changes occurring at different levels to be judged. The importance of valuation applies to each of the levels of the framework by Fineberg and co-workers. For example, if patients experience positive changes in diagnostic accuracy (e.g. greater diagnostic certainty) but negative changes in therapeutic impact (e.g. additional invasive procedures) following the use of a new imaging technology, has a net benefit been provided overall? In order to answer such questions, an economic evaluation must estimate values for the range of possible consequences. From the review of studies described in the next chapter, it would be appropriate to describe most economic evaluations of diagnostic imaging technologies as cost–consequences analyses, since they provide a description of a range of benefit parameters and do not attempt to aggregate or value the different components. A number of methods exist by which the estimation of values, or preferences, for a range of different outcomes can be undertaken. They are collectively referred to as stated preference techniques and include contingent valuation and conjoint measurement. The feasibility of using such approaches to value the range of benefits associated with diagnostic imaging technologies was explored in this research.

Chapter 3

The role of MRI in the diagnosis of knee injuries: a brief review

Clinical history and findings at physical examination in patients with abnormalities of the knee are known to be non-specific in the determination of the cause of the abnormality.[12] In patients in whom the diagnosis is uncertain, physicians must, therefore, turn to other diagnostic modalities to inform the selection of appropriate treatments. Arthrography has the potential of a high level of accuracy, but is invasive and technically very demanding. For these and other reasons, arthroscopy has become the diagnostic procedure of choice. Although diagnostic arthroscopy is an invasive and relatively high-cost procedure, proponents point to its accuracy and to the surgeon's ability to diagnose and treat abnormalities with a single intervention. However, diagnostic arthroscopy sometimes reveals no abnormality or only minor lesions. Thus, some patients may be subjected to a surgical procedure, with its associated risks, without the reward of treatment or cure. Thus, orthopaedic surgeons are increasingly turning to MRI as a non-invasive means of diagnosing knee problems. Whilst MRI appears sensitive for the diagnosis of medial meniscal tears and injuries to the anterior cruciate ligament (ACL),[13,14] it has lower sensitivity for injuries to the lateral meniscus, medial collateral ligament, patella retinaculum and articular cartilage.[15] Furthermore, the specificity and sensitivity of MRI decreases as the number of injured structures within the knee increases[16] and in the presence of a heamarthrosis.[15] Specificity and sensitivity are also dependent on the reporting radiologist and the strength of the magnetic field.[14] The fact that an MRI scan of the knee shows an effusion only does not exclude a significant injury.[15] Conversely, an abnormal MRI finding may not necessarily be the cause of a patients symptoms, since up to 13% of asymptomatic patients under 45 years of age may have a meniscal tear on MRI.[17]

Much of the technology assessment work relating to MRI has focused on its use in the neurosciences for imaging of the head and spine.[18,19] The use of MRI in orthopaedics has not been extensively evaluated and the work that has been done has tended to focus on the narrow issue of accuracy and not explored costs and benefits more widely.[14,20–27] For example, the sensitivity and specificity of 0.5 Tesla MRI in diagnosing internal derangements of the knee in one patient series has been compared to published data on 1.5 Tesla MRI.[21] Warwick and colleagues[28] and Franken and colleagues[29] are two rare studies that investigated the impact of MRI on clinical management.

Warwick and colleagues[28] reported data on the diagnostic accuracy of MRI for patients that presented with knee injuries from an observational study. Patients on the waiting list for diagnostic knee surgery were offered an MRI scan and on the basis of the results of the scan 32% were removed from the list, since their injuries did not require surgical repair. The sensitivity of MRI for this patient group was found to be 100% and the specificity was 66.7%. However, the data on diagnostic accuracy were based on the subsample of patients who went on to receive surgery, since findings at surgery were taken as the 'gold standard' diagnosis.

A fairly typical example of an evaluation of a diagnostic imaging technology was performed by MacKenzie and co-workers[30,31] who investigated the effectiveness of MRI of the knee. They observed a single cohort of patients and measured the diagnostic impact of MRI in terms of changes in diagnosis and diagnostic confidence, as judged by clinicians before and after the imaging examination. Diagnoses were specified in terms of the probable anatomical site of the lesion under investigation and the working diagnosis for that lesion. Clinician confidence was measured using a visual analogue scale (VAS). A total of 324 patients were included in the study. MRI was found to have a diagnostic impact in some cases by refuting certain clinical diagnoses and in others by improving clinician confidence in diagnosis. Additionally, the use of MRI helped to bring about new, previously unsuspected diagnoses in 21% of patients. In terms of changes in the diagnostic and treatment pathways, there was a marked shift away from the use of surgery either for diagnosis or treatment.

Thus, the findings of the study suggested that MRI significantly influenced clinicians' diagnoses and management plans in patients with knee problems.

Hollingworth and colleagues[32] reported an investigation of changes in patient quality of life following MRI of the knee. This study was not a full economic evaluation, but provides data that might be useful for the purpose of such evaluation. Health-related quality of life was measured using a number of instruments, including the Short Form with 36 items (SF-36) and the EQ-5D. Health changes were measured in patients referred for MRI of the knee using a baseline assessment, undertaken before the imaging examination, and a follow-up survey at 6 months. The SF-36 and EQ-5D instruments indicated statistically significant ($p < 0.05$) improvements in quality of life at 6 months, in general, although the patients remained at levels below the general population norms. For the SF-36, significant improvements were recorded on five of the eight dimensions: physical functioning, role limitations (physical), bodily pain, social functioning and mental health.

Whilst many studies have investigated the sensitivity and specificity of MRI for the detection of internal derangement of the knee,[27,33–36] and others have examined the role of MRI in improving diagnostic accuracy and reducing the need for diagnostic arthroscopy,[12,37–39] there are relatively few studies exploring the role of MRI in the management of the acutely injured knee. Those papers that have addressed this question have reached conflicting conclusions.[40–42] Furthermore, the effect of MRI availability on the management process and the final functional outcome after an acute (or even chronic) knee injury has rarely been evaluated.[12,30,43] In addition, no study has used data from an RCT; the majority of studies have been based on case-series and, thus, the results must be interpreted cautiously. Given the nature of the existing literature, it is, therefore, still an open question whether the use of MRI for knee investigations represents a cost-effective diagnostic procedure.

Chapter 4

The role of MRI in the diagnosis of knee injuries: the Kent and Canterbury trial

Introduction

This chapter describes an empirical investigation of the use of the diagnostic imaging technology MRI in the diagnosis of knee abnormalities and injuries in a DGH setting. The principal purpose of this research was to determine whether, for patients presenting in a DGH with a knee problem who were being considered for surgery, MRI has a major impact on:

- clinical management
- health sector and patient costs
- patient outcome.

The research was based on a single-centre RCT conducted at Kent and Canterbury Hospital, Kent, UK. Research Ethics Committee approval was obtained prior to the commencement of the study.

Methods

Patient populations

All study patients were recruited from routine orthopaedic clinics, which they attended following referral from either their general practitioner (GP) or the Accident and Emergency Department. The aim was for trial patients to be representative of the range of knee problems seen in orthopaedic outpatient clinics at a DGH. Eligibility for inclusion in the trial was assessed for all patients with a knee problem who attended the orthopaedic clinics run by participating surgeons. Recruitment was undertaken between February 1996 and August 1997. The patient information sheet and consent form used in this study are provided in appendix 1.

Patients were defined as suitable for the trial if:

- diagnostic or therapeutic arthroscopy was being considered (in the absence of MRI)
- there had been no previous major surgery in the injured knee, such as knee replacement (previous arthroscopy and partial menisectomy did not exclude patients from the trial)
- there was no pre-existing chronic knee pathology
- there was no serious condition requiring immediate attention, e.g. a serious knee infection
- there was no history or current experience of recurrent locking of the knee
- they were aged between 16 and 55 years
- anterior knee pain was not the main clinical indication.

Details of the sample size calculations are provided in appendix 2.

Research process and treatment pathways

Once written consent was obtained, study patients were allocated to one of two trial arms: investigation using an MRI scan (MRI trial arm) and investigation using arthroscopy (no-MRI trial arm). The on-site study researcher allocated patients to one of the trial arms using randomly ordered opaque sealed envelopes. Patients allocated to the MRI trial arm were booked for an MRI scan (median wait for a scan = 29 days) and placed immediately on the arthroscopy waiting list, even though they may not have required surgical treatment. This management pathway was adopted on the advice of the Hospital Research Ethics Committee in order to ensure that patients were not 'disadvantaged' as a result of participating in the trial. Patients were reviewed in an outpatient clinic following their scan and a decision on appropriate management was made. Routine clinical follow-up then continued until the knee had recovered. Patients allocated to the no-MRI trial arm were immediately listed for arthroscopy, reviewed in clinic (both before and after surgery) and followed up according to routine clinical practice until the knee problem resolved.

It was clearly neither feasible nor sensible to blind the study patients, researchers or those involved in providing care to the outcome of the allocation process and, thus, an open-label policy was adopted.

Assessment of benefits

Using the categorisation by Fineberg and colleagues,[4] this study investigated the benefits

of knee MRI at two levels: diagnostic/therapeutic impact and patient outcome. The use of arthroscopy as a form of diagnosis as well as a therapeutic intervention complicates the picture for knee investigations and prevents a complete separation in this study of diagnostic and therapeutic impacts. The observed data on whether or not surgery was undertaken over the 12-month follow-up period allow an assessment of the diagnostic and therapeutic impact of MRI. From a patient's perspective, diagnostic or therapeutic arthroscopies are virtually identical: both are typically day-case procedures involving similar recovery periods.

Benefits in terms of patient outcome were measured using two generic health-status measurement instruments, the SF-36[44–47] and the EQ-5D.[48–50]

The SF-36 has eight dimensions: physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role-emotional and mental health. The dimensions relating to physical functioning, physical role limitations and bodily pain were thought to be of particular relevance for this patient group. The physical functioning dimension addresses the extent to which health limits physical activities such as self-care, walking, climbing stairs, bending, lifting and moderate and vigorous activities. The physical role limitation dimension examines the extent to which physical health interferes with work or other daily activities, and the bodily pain dimension establishes the intensity of any pain and the effect of pain on normal activities. For each dimension, a score can be calculated (range 0–100), which provides a broad indication of severity (higher scores indicate less severe states). The standard scoring rules for the UK version of SF-36 were used giving a score for each dimension of the instrument, allowing comparison of patients' quality of life between trial arms. Missing data were coded in line with standard coding rules. This involved prediction of missing responses within dimensions, but only in cases where more than 50% of questions relating to the dimension had been answered.

The components of the EQ-5D instrument used here were those seen as standard for use in clinical trials: the EQ-5D health-state categorisation and the health-state 'thermometer' or VAS. The EQ-5D has five dimensions (mobility, usual activities, self-care, pain/discomfort and anxiety/depression), each with three levels (no problems, some problems and extreme problems). This allows 243 unique health states to be defined. For each state, there exists a tariff value obtained from a survey of a large sample of the population of England and Wales, using the time trade-off technique.[51] The EQ-5D VAS component requires respondents to indicate the severity of their current health state on a scale ranging from 0 to 100 (anchored by 'worst imaginable health state' and 'best imaginable health state'). The advantage associated with using the EQ-5D is that overall quality of life can be compared between groups. This is currently not possible using the SF-36, since comparison takes place within each dimension. Thus, if improvement is seen on one dimension (e.g. physical functioning) but deterioration occurs on another (e.g. mental health), the SF-36 cannot indicate whether the patient is in a preferred health state overall.

Assessments of quality of life were taken at baseline (at the point of recruitment) and at 6 and 12 months after recruitment into the trial. In most cases, the baseline questionnaire was completed at the clinic visit. When the patient was unwilling to make an immediate decision on their participation in the trial, they took the study information home and, for those who chose to take part, the baseline questionnaire was sent by post. The two follow-up questionnaires (at 6 and 12 months) were both distributed by post with pre-paid return envelopes. Non-responders were sent two reminder letters with a further copy of the questionnaire. All questionnaires were returned to Brunel University in order to overcome potential Hawthorne effects.

Assessment of costs

The perspective for the cost analysis reported here was the NHS and patients, such that the main objective was to identify all important NHS and patient resource use and cost differences between trial arms. The first step in the costing process was to collect data on the NHS resources devoted to, or consumed by, trial patients. Data were collected on the following resource parameters: the MRI scan, the surgery on the knee, other relevant diagnostic procedures, other relevant therapeutic procedures, relevant drugs prescribed, associated outpatient attendances, associated inpatient episodes and relevant contacts with other healthcare professionals, such as physiotherapists and GPs. Most of these data were extracted from patient medical records using a process of case review by the on-site study researcher, which continued for up to 12 months after recruitment. Data on the use of community healthcare resources, such as GPs or community physiotherapists, were collected using the patient

follow-up questionnaires distributed at 6 and 12 months. Some resource-use data were missing where patients failed to return the relevant questionnaire. However, given that the major aspects of NHS resource use (i.e. hospital-based resource consumption) are not censored for any trial patients, a process of mean substitution for missing resource-use data was undertaken using groups defined by age, sex and trial arm.

The healthcare resource-use items were costed using unit cost information from three sources (see *Table 1*). Unit costs for MRI, arthroscopy, orthopaedic outpatient visits and orthopaedic physiotherapy sessions were obtained from a survey, undertaken as part of this project, of ten hospital Trusts that provided such services. The unit costs in the costing exercise were the mean costs across the ten Trusts. In order to provide verification of the unit cost estimates for arthroscopy, a separate time-and-motion study was undertaken in order to collect data on the resources used in ten arthroscopy procedures carried out by orthopaedic surgeons from Kent and Canterbury Hospital. This allowed a bottom-up estimate of cost to be derived. Given that this was very similar to the mean cost derived from the survey of ten Trusts, the survey-based estimate was used in the analysis presented here. For GP visits, unit cost information was obtained from the University of Kent annual survey of costs of community health services,[52] and the unit cost of X-rays were obtained from University Hospital Birmingham. All costs were brought to a common 1998 price base.

The costs incurred by patients in attending for an MRI scan, outpatient visit or inpatient/day-case procedure were also investigated through a survey of a subsample of all trial patients. Selected patients were asked to complete a simple questionnaire detailing their transportation costs, time costs (travelling time and time at the hospital) and other self-reported costs. Patients were also asked whether or not they had a companion with them and if that companion had an appointment in the hospital/primary care clinic. Companion costs were excluded from this study if the companion also had an appointment at the hospital.

Under the heading of 'transportation costs', data were collected on the mode of transport used in travelling to the hospital. If the patient travelled by car they were asked to estimate the approximate distance travelled (in miles), which was multiplied by an average motoring cost per mile of £0.41, including depreciation and running costs.[53] If the patient travelled by taxi or public transport they were asked to report their normal fare. Data on the use of, and associated out-of-pocket costs associated with, any other form of transport (e.g. ambulance, walking, cycling) were also collected. The total travel cost incurred by each patient (and companion) in the subsample was then estimated. For the estimation of 'time costs', patients were asked to estimate the time involved in their journey door-to-door and the time spent at the hospital. In addition, patients provided an indication of what they and their companion would otherwise normally have been doing. When work time had been given up, the time estimate was costed using an average wage rate of £9.54 per hour.[54] Non-working time (leisure time) was valued at £3.41 per hour, following previous methods adopted on this issue which advise that all non-working time should be valued regardless of the age of the individual.[55,56] Total patient costs were calculated by summing costs associated with travel, time and other expenses.

Therefore, patient-specific resource-use data (both health sector and private resources) were available on a per-patient basis. Using the relevant unit costs

TABLE 1 *Unit cost estimates*

NHS resources	Unit cost (£)[a]	Source
Arthroscopy procedure	485.00	Survey of ten NHS Trusts
ACL repair	2194.00	Survey of ten NHS Trusts
First outpatient visit	89.00	Survey of ten NHS Trusts
Subsequent outpatient visit	44.50	Survey of ten NHS Trusts
Knee MRI scan	138.50	Survey of ten NHS Trusts
Knee X-ray	25.00	University Hospital Birmingham
Physiotherapy session	31.00	Survey of ten NHS Trusts
GP visit	15.00	Netten and Dennet, 1998

[a] *1998 prices*

for each resource item, estimates of the health sector and total (i.e. health sector plus patient) cost per patient was generated.

Data analyses

All analyses were conducted on the basis of 'intention-to-treat'. Each aspect of analysis is detailed below.

Comparisons of baseline characteristics, including quality of life

For the between-group comparisons of baseline patient characteristics and baseline quality of life, the following analyses were conducted as a test of the success of randomisation. Where comparisons involved continuous data, either parametric (student's *t*-test) or non-parametric (Mann–Whitney Wilcoxon test) tests were used, depending on the degree of deviation from normality in the sample distributions. Where the comparisons were of proportions, either a Π^2 test or, where the number in one or more cells was small (i.e. frequency < 5), Fisher's exact test was used.[57]

Descriptive analyses of changes in quality of life over time

For all quality-of-life scores (all eight SF-36 dimension scores and the EQ-5D VAS and tariff scores), the profile over time was plotted for every study patient, including those for whom the profile was not complete due to non-response. Whilst such plots provide a visual impression of whether between-group differences exist in profiles, they do not, in themselves, allow estimates and associated statistical significance of any differences to be determined.

Comparison of quality of life at each time point

At each follow-up time point (i.e. at both 6 and 12 months), between-group comparisons were made using the same methods as adopted in the baseline comparison described above. The main limitations of this approach are that it takes no account of the longitudinal nature of the data, involves multiple testing and ignores problems of sample selectivity due to non-response.

Longitudinal analysis of quality of life using summary measures

Mean changes in quality-of-life scores were estimated separately for the 0–6-month change, the 0–12-month change and the 6–12-month change, using a 'complete case analysis'. The between-group comparisons of such changes were undertaken using the same methods as adopted in the baseline comparison described above. The main limitation of presenting data on changes in this manner is that patients who have provided only partial data (e.g. where a patient has responded to two but not all three questionnaires) are omitted from some of the comparisons. This may be a source of bias, especially where the response rate is low. In addition, this approach again involves multiple testing.

Econometric analyses

In order to allow for potential differences between groups at baseline, quality-of-life scores were also compared using regression analyses. A two-limit tobit model with random effects and sample selectivity was developed as part of the research reported here (technical details of this new technique are provided in appendix 3). Models were estimated for the six SF-36 dimensions for which data can be treated as continuous (i.e. physical functioning, bodily pain, general health, vitality, social functioning and mental health) and the EQ-5D tariff and VAS scores.[58] These models take into account the fact that the dependent variable is limited in range (i.e. SF-36 and EQ-5D VAS scores cannot be outside the 0–100 range, and EQ-5D tariff scores are bounded by –0.594 and 1) although they can be treated as continuous within that range. The independent variables included in the models were trial arm, hospital site, referral source, patient sex and patient age at baseline. The random effects allow the data collected at each time point to be treated as a separate panel, while controlling for possible correlation between disturbances over time due to time-persistent unobserved heterogeneity. The sample selection component attempts to control for any bias due to non-responders at both the 6- and 12-month follow-up surveys, and shows the extent to which individual characteristics such as sex, age and trial arm allocation can be used to predict whether or not an individual will participate at 6 and 12 months.

Cost analyses

The analysis of cost data followed the guidance provided by Briggs and colleagues.[59] Firstly, the distributional form of the data was investigated separately for the two arms of the trial in order to identify the extent of any skew that might present problems for standard parametric statistical tests. This involved plotting the frequency distributions of the per-patient cost data and visual inspection of these distributions. Secondly, since skewed distributions for cost data are commonly found, and were expected a priori in this study, a non-parametric approach to the cost analysis was likely to be required. This was particularly the case here given the relatively small sample size, which may

not have been sufficient for the assumption of a normal sampling distribution (from the central limit theorem) to be justified. However, estimates of the mean cost were appropriate for use in the economic evaluation, even where the data were skewed, since we were interested in both the average per-patient cost of a particular treatment and (because there exists a budget constraint) the total cost of care for a patient group. Thus, the non-parametric approach of bootstrapping was used to estimate confidence limits around the estimates of mean cost.[59,60]

Economic evaluation

Framework

Data are presented below on the full costs and benefits associated with the management patterns adopted in the two arms of the trial. For the economic evaluation, an incremental analysis was conducted with a focus on the additional costs and additional benefits associated with patients receiving an MRI scan. The first component of the economic evaluation was a cost–consequences analysis in order to assess whether one strategy was dominant (i.e. associated with both a lower cost and an unambiguously better outcome on all dimensions). It is uncommon to find dominance and further analyses are usually required that make explicit the nature of the trade-off between costs and benefits. The data collected for this study allowed for the possibility of cost–utility analysis to be undertaken, since quality-adjusted life-years gained could be estimated using the EQ-5D tariff scores.[51]

Dealing with uncertainties

As with any economic evaluation, there were uncertainties associated with the data and the analyses presented in this chapter. The data inputs themselves were uncertain and this was dealt with largely by using the statistical methods described above. Uncertainty also existed in the cost of MRI scans and, thus, the one data input upon which sensitivity analysis was conducted was the unit cost of an MRI scan. Recent technological advances have seen the development of smaller MRI units for use in the imaging of extremities, including the knee joint. The base-case analysis assumed the use of a traditional full-scale MRI installation, but the sensitivity of the results to the consideration of an extremity-specific scanner was explored using one-way sensitivity analysis as reported below.

As indicated in chapter 2, the importance of positive consequences associated with the use of MRI in terms of diagnostic and therapeutic impacts depends on the strength of the preference that the patients have for such changes. If the trial data indicated that MRI was associated with the avoidance of surgery in some patients, then uncertainty exists in terms of the importance or value to place on this finding. This aspect of uncertainty is explored empirically in chapter 7.

Results

Figure 2 shows the trial profile, indicating the number of patients recruited into the study, the randomisation assignment, the numbers receiving an MRI scan and the number of measurements for each randomised group.

Sample characteristics and comparability of groups

A total of 118 patients consented to take part in the trial. Of those, 59 patients were allocated to the MRI arm and 59 to the no-MRI arm. At baseline, the two groups were well matched in terms of:

- age (MRI group: mean = 36 years, range 16–55; no-MRI group: mean 36 years, range 17–54; $p = 0.86$),
- sex (proportion of females in MRI group = 32%, proportion of females in no-MRI group = 37%; $p = 0.56$),
- referral source (proportion of GP referrals in MRI group = 88%, proportion of GP referrals in no-MRI group = 91%; $p = 0.40$),
- injured leg (proportion of injuries of right leg in MRI group = 49%, proportion of injuries of right leg in no-MRI group = 52%; $p = 0.93$),
- duration of knee problem (median in MRI group = 28 weeks, median in no-MRI group = 36 weeks; $p = 0.40$).

The overall response rate to the baseline quality-of-life questionnaire was 91.5% (overall = 108/118, MRI group = 57/59, no-MRI group = 51/59). The baseline comparisons of quality-of-life scores are shown in *Figure 3*. Given that data on all SF-36 dimensions and the EQ-5D VAS and tariff had approximately normal distributions, means and 95% confidence intervals (CIs) are presented. There were no statistically significant differences between groups for any of the SF-36 dimensions at baseline (*t*-test minimum $p = 0.10$ for all dimensions). In addition, no trend of between-group differences at baseline was evident: the no-MRI group had a higher score on five of the eight SF-36 dimensions. The data indicate that at the point of recruitment into the trial, patients tended to have particular problems relating to three SF-36 dimensions: bodily pain, role-physical and vitality. The role-physical dimension is principally

concerned with the extent to which health problems limited work or other principal daily activities.

Similarly, both the EQ-5D tariff ($p = 0.72$) and VAS ($p = 0.91$) scores were not statistically significantly different between groups at baseline. *Figure 4* shows the baseline data on the percentage of respondents who indicated that they had no problems, some problems or extreme problems for the five dimensions covered by EQ-5D. In line with the SF-36 data, the principal problems were related to mobility, usual activities and pain and discomfort, and no statistically significant between-group differences were found (comparison of proportions of respondents with no problems: minimum $p = 0.10$ for all dimensions).

Diagnostic/therapeutic paths

A total of 71 operative procedures were undertaken over the course of the 12-month follow-up period, on 66 study patients (55.9% of all trial patients). There was a statistically significant difference between groups in terms of the proportion of patients who received surgery (no-MRI arm = 0.71; MRI arm = 0.41; $p = 0.001$). The majority of operative procedures were arthroscopies (90.1% of all procedures), most of which were undertaken on a day-case basis. One patient in each trial arm had two arthroscopies in the 12-month follow-up period. The other operative procedure undertaken in the 12-month follow-up period on some study patients was acute repair of the ACL. The use of this procedure was evenly distributed between trial arms: in the 12-month follow-up period, two patients each had a single ACL repair procedure in the MRI arm and one patient underwent two ACL procedures in the no-MRI arm.

For those patients who received surgery, the time interval from randomisation to surgery was not significantly different between the two groups (no-MRI arm: mean = 152 days, standard deviation (SD) = 106; MRI group: mean = 173 days, SD = 119; $p = 0.47$). This result is an artefact resulting from the requirement of the hospital's Ethics Committee that patients allocated to the MRI arm were put on the arthroscopy waiting list at randomisation in order for them not to be disadvantaged as a result of participating in the trial. Thus, it would be inappropriate to assume that this specific result would be seen in another setting, unless a similar policy of listing for surgery was in operation. The mean (SD) time from randomisation to receiving an MRI scan was 42 (35) days, with a range of 7–141 days. A total of nine patients in the MRI arm did not attend for their MRI scan.

Response rates and sample selectivity

The overall response rates to the follow-up quality-of-life questionnaires were 67.8% at 6 months (overall = 80/118, MRI group = 44/59, no-MRI group = 36/59) and 58.5% at 12 months (overall = 69/118, MRI group = 40/59, no-MRI group = 29/59). For the 12-month follow-up, the higher response from the MRI group was statistically significant (difference between proportions = 0.186; 95% CI, 0.012 to 0.361).

The extent to which further bias was introduced through non-response was also investigated for the follow-up data by comparing the baseline characteristics of responders and non-responders. As indicated below, at 6 months the two groups were generally well matched:

- age (mean for non-responders = 33 years, mean for responders = 36 years; $p = 0.08$)
- sex (proportion of females in non-responder group = 39%, proportion of females in responder group = 32%; $p = 0.46$)
- referral source (proportion of GP referrals in non-responder group = 81%, proportion of GP referrals in responder group = 93%; $p = 0.06$)
- duration of problem (median for non-responders = 20 weeks, median for responders = 32 weeks; $p = 0.09$)
- injured leg (proportion of injuries of right leg in non-responder group = 39%, proportion of injuries of right leg in responder group = 56%; $p = 0.12$)

There was a tendency, however, for non-responders to be slightly younger and to have had their knee problem for a shorter period, but none of the differences reached conventional levels of statistical significance. In addition, the comparison of the two groups (non-responders versus responders) in terms of their baseline quality-of-life scores showed no statistically significant differences between groups for any of the quality-of-life dimensions (minimum $p = 0.18$ for all dimensions) and there was no strong trend in favour of one group.

Non-responders and responders were less well matched at 12 months: non-responders were significantly younger and had had their knee problem for a significantly shorter period of time. Whilst the differences between responders and non-responders reached statistical significance for these two baseline characteristics, the groups were similar in all other respects:

- age (mean for non-responders = 33 years, mean for responders = 37 years; $p = 0.03$)

- sex (proportion of females in non-responder group = 39%, proportion of females in responder group = 32%; $p = 0.44$)
- referral source (proportion of GP referrals in non-responder group = 86%, proportion of GP referrals in responder group = 93%; $p = 0.26$)
- duration of problem (median for non-responders = 20 weeks, median for responders = 36 weeks; $p = 0.03$)
- injured leg (proportion of injuries of right leg in non-responder group = 45%, proportion of injuries of right leg in responder group = 55%; $p = 0.49$)

The comparison of baseline quality-of-life scores between the non-responders to the 12-month survey and the responders revealed that, whilst only one of the quality-of-life dimension scores reached statistical significance (i.e. EQ-5D thermometer, $p = 0.02$), there appeared to be a trend for responders to have had higher quality-of-life scores at baseline. Thus, the general finding was that some bias may have existed in the data from the 12-month follow-up survey due to the disappointing response rate, and, thus, the 12-month data should be interpreted with caution.

Benefits

Descriptive analyses of changes in quality of life over time

To examine the nature of the quality-of-life changes taking place over time, the individual quality-of-life profiles were plotted for every study patient (even those for whom the profiles were not complete) for all quality-of-life dimensions. From visual inspection of these individual profiles, there did not appear to be any clear patterns of change that distinguished between the MRI and no-MRI groups.

Comparison of quality of life at each time point

The comparisons of quality-of-life scores at 6 months after recruitment into the trial are shown in *Figure 5*. Again, the data on all SF-36 dimensions and the EQ-5D VAS and tariff had an approximately normal distribution and, thus, means and 95% CIs are presented. There were no statistically significant differences between groups for any of the SF-36 dimensions at 6 months (*t*-test minimum $p = 0.09$ for all dimensions). However, a trend of between-group differences is evident because the mean scores for the no-MRI group are higher (or the same) across all but one SF-36 dimension (role-emotional). These data also suggest that, for many patients in both groups, problems continued in terms of bodily pain, performing daily work or other activities (role-physical) and vitality. Similarly, both the EQ-5D tariff and VAS scores were not significantly different between groups at 6 months (*t*-test $p = 0.13$ and $p = 0.88$, respectively). *Figure 6* shows the 6-month data for the five EQ-5D dimensions. In line with the SF-36 data, the principal problems reported at 6 months continued to concern mobility, usual activities and pain and discomfort.

The pattern in the quality-of-life data seen at 6 months was repeated at 12 months. The means and CIs are presented in *Figure 7* (approximately normal distributions for all dimensions) and indicate a trend of between-group differences: the mean scores for the no-MRI group are higher across all parameters. *Figure 8* shows the 12-month data for the EQ-5D dimensions, which indicate continuing problems for some patients in all dimensions other than self-care.

Longitudinal analysis of quality of life using summary measures

Figures 9 and *10* show the change in quality-of-life scores (both SF-36 and EQ-5D) from 0 to 6 months and from 0 to 12 months, respectively. By definition, these figures only report data for study patients who returned both the baseline and the relevant follow-up questionnaires. A similar trend was seen in all dimensions at 12-month follow-up in favour of the no-MRI group (see *Figure 10*): larger positive changes (or smaller negative changes) in quality-of-life scores were seen for all SF-36 dimensions and for the two EQ-5D dimensions. However, as mentioned above, the main limitation of such analyses was that patients were excluded if they did not respond to all three questionnaires.

Econometric analyses

The results of the two-limit tobit models with random effects and sample selectivity estimated for each SF-36 dimension and the EQ-5D tariff and VAS scores are reported in *Tables 2–9* and described below.

As confirmation of the success of randomisation, there were no statistically significant differences between trial arms in any of the quality-of-life dimensions at baseline.

The sample selection models identified characteristics that predicted participation in both the 6- and 12-month follow-up surveys. Whilst there was some variation between quality-of-life dimensions, in general, response appeared to have been associated with patient age (i.e. younger patients were less likely to respond) and trial

arm (i.e. patients allocated to the no-MRI arm were less likely to respond).

The main models estimated using data from the 6- and 12-month surveys and adjusted for sample selectivity showed no strong evidence of differences between trial arms. The dummy variable for trial arm failed to reach conventional levels of statistical significance in all models. However, the trial arm variable had a negative coefficient in virtually all follow-up models (at both 6 and 12 months). This finding was consistent with a trend towards higher quality-of-life scores for patients in the no-MRI group discovered in the earlier analyses.

Resource use and costs

Summary information on NHS resource use and costs, broken down by trial arm, is presented in *Table 10*. A similar pattern of resource use was found in the two groups in terms of outpatient attendances, drug costs (excluding drugs used in surgery), physiotherapy sessions and GP visits. Other than MRI, very few further investigations were undertaken on study patients. As indicated above, the difference between trial arms in the proportion of patients who received surgery was statistically significant.

Table 10 also reports estimates of the NHS cost per patient. The frequency distributions for NHS costs by trial arm, using data from both sites, are shown in *Figures 11* and *12*. The data clearly had a very skewed distribution: in both groups, some patients were associated with a relatively low cost and some with a relatively high cost, which was, in part, dependent on whether they received surgery. No statistically significant difference was found between mean costs (i.e. the bootstrap 95% CI for the difference between groups crossed zero as shown in *Table 10*). Thus, overall, the mean NHS cost per patient was similar for patient management with and without MRI.

Table 10 also reports results for patient costs and for both NHS and patient costs combined. The mean patient cost was a little higher in the MRI group. However, this was unsurprising given that virtually all patients in that group had the additional travel and time costs associated with attending for the MRI scan. When all costs were considered (both NHS and patient costs) the results mirror those for NHS costs only with no statistically significant difference.

On the basis of the list price and discussions with a manufacturer of extremity-specific MRI equipment, we assumed that the unit cost for an MRI scan of the knee produced using an extremity-specific MRI scanner would be one-third of the price of a scan produced using a conventional MRI installation. The cost data were re-analysed using the lower unit cost for an MRI scan of £46. As expected, the results indicated a lower mean cost for the MRI group, but the findings were very similar to those found in the base-case analysis overall: no statistically significant difference in mean NHS costs between groups (MRI group: mean = £679, SD = 804; no-MRI group: mean = £703, SD = 602).

Economic evaluation

The results from this research were brought together in the form of a cost–consequences economic analysis under the following headings: patient health outcome, diagnostic and therapeutic impacts and costs.

When patient health outcomes were judged from a traditional health economics perspective, the central finding was that there was no significant difference between groups in health-related utility (measured using the EQ-5D tariff scores). Similarly, on all other measures of health outcome (i.e. EQ-5D VAS and all SF-36 dimensions) no statistically significant differences were revealed between groups, although a trend in favour of the no-MRI group was observed.

However, the use of MRI was found to be associated with a positive diagnostic and therapeutic impact: a significantly smaller proportion of patients in the MRI group underwent surgery in the 12-month follow-up period (41% of MRI patients had surgery compared to 71% of no-MRI patients). This suggested that, where the alternative diagnostic strategy is often arthroscopy, the use of MRI potentially has positive benefits for patients with chronic knee problems through the avoidance of unnecessary surgery.

The data suggested a similar mean cost for the NHS for both the MRI and no-MRI groups. On average, the additional cost associated with providing MRI scans to all patients was offset by the reduction in the proportion of patients who underwent surgery.

Therefore, the cost–consequences analysis identified a potentially dominant technology overall: whilst the use of MRI was not found to be associated with improved health outcomes or reduced costs, a positive impact was demonstrated in terms of the avoidance of surgery. Given that

no differences in either health utility or cost were revealed in this study, it is neither possible nor sensible to construct a traditional cost–utility ratio. The magnitude of the benefit seen in avoiding surgery, in some cases, depended on the strength of preference for such 'in process' outcomes. This issue is explored further in chapter 7.

Discussion

The principal finding of this research was that the use of MRI in patients presenting principally with chronic knee problems had a positive diagnostic/therapeutic impact in reducing the risk of surgery.* This represents a potential benefit to patients given that surgery involves inconvenience (i.e. 1 day, or sometimes longer, in hospital and typically several further days until full recovery) and risks associated with both the use of anaesthesia and the surgery itself. Whilst such risks are low in frequency and, hence, were not observed in the trial cohort, they can be very serious in terms of both morbidity and mortality. Therefore, in this study, the reduced exposure to such risks represents an unquantified but important benefit to patients. The additional cost associated with providing MRI to all patients was offset in full through the avoided costs of surgery in some patients, which made it cost-neutral. Other benefits associated with MRI in this patient population, such as improved outcomes or enhanced health-related quality of life, were not observed.

One of the disappointing features of the data collected as part of this trial was the relatively high non-response rate to the follow-up questionnaires at 6 and, especially, 12 months. One possible explanation for this relates to the age group of the sample. Firstly, it is well known that questionnaire response is related to age, with younger people generally being poorer responders, and secondly, younger people are more likely to be mobile and, thus, some study patients may have moved house during the course of the study. Interestingly, the level of non-response at 12 months was higher in the no-MRI group. One interpretation of this finding could be that patients may have been disappointed at not having been allocated to the MRI group. At the time of recruitment, all patients were told about the potential benefits of MRI as a new form of diagnosis, but those allocated to the no-MRI group were then denied this new alternative and, as a result, may have been less inclined to respond to the questionnaire. As described earlier, the extent to which the poor response introduced bias was investigated and, in general, the non-responders at 12 months tended to be those with lower (i.e. poorer) quality-of-life scores at baseline. Therefore, since the responders represented a group with a better quality of life, the lower response in the no-MRI group would suggest a better quality-of-life picture, on average, for those that did respond. This is the finding that emerged from the trial and whilst one of the analytical methods used (i.e. the two-limit censored regression models with random effects and sample selectivity) adjusted for non-response, some caution is required in the interpretation of the quality-of-life results.

Another limitation of the study concerns the follow-up period of 12 months. The consequence of this is that all patients were censored at that point, even if they had not been censored earlier. This may have had a particularly important impact on some of the key resource-use parameters, in particular the use of arthroscopy. The time from randomisation to surgery amongst all study patients ranged from 6 to 352 days. The fact that some study patients underwent surgery towards the end of the 12-month follow-up period suggests that a longer follow-up period would almost certainly have seen a larger number of patients receiving surgery. However, given the necessity of the trial design that all patients were listed for surgery regardless of the trial arm to which they were allocated, it is not certain that a longer follow-up would have necessarily caused the difference in the proportions of patients receiving surgery to become less between the arms.

Finally, the use of a cost–consequences framework leaves unanswered the central question of whether the expected gain associated with the positive consequence (the diagnostic/therapeutic impact of avoiding surgery) justifies the use of MRI in patients with chronic knee problems. One response to this is that decision-makers should be presented with the evidence, allowing the trade-offs inherent in the resource allocation problem to be clear, so that the responsibility for making the decision is then in their hands. However, the implicit assumption with this response is that the decision-makers' values are those that are appropriate. If the alternative view is taken that the allocation of public

* A similar trial concerned with the use of MRI in patients presenting with acute knee injuries is reported in appendix 4.

resources should take into account the values of the potential patients or citizens then more than just the cost–consequences framework may be required. In addition, since the consequences being discussed here relate to short-term outcomes linked to 'process', especially those concerned with diagnostic and therapeutic impacts, the traditional cost–utility framework with an emphasis on measuring utility in terms of quality-adjusted life-years may also be inappropriate. Therefore, for the evaluation of diagnostic imaging technologies, alternative forms of analyses may sometimes be required to enable the estimation of the value associated with the full range of consequences. This issue is discussed further in chapter 7.

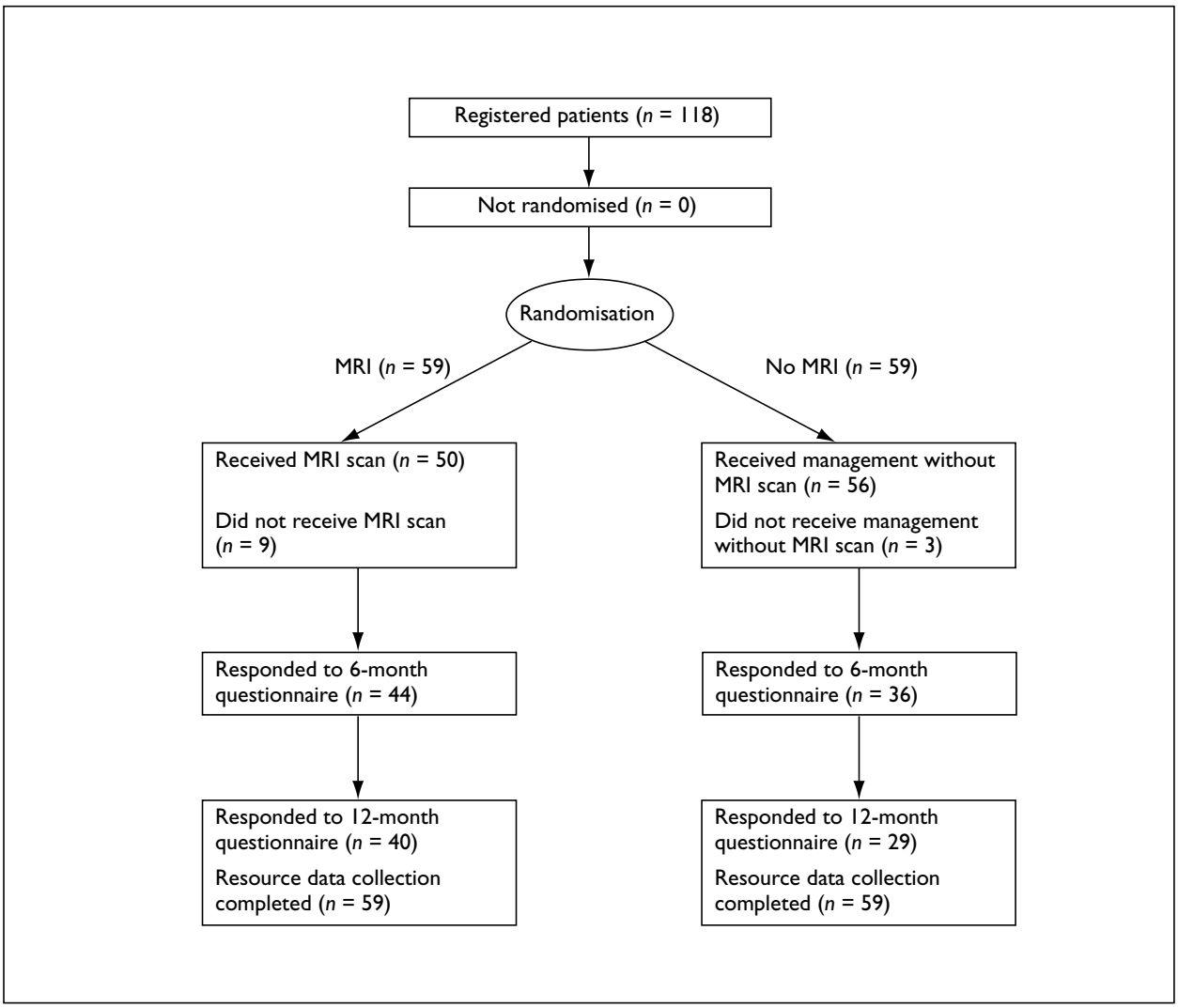


FIGURE 2 *Flow chart describing the progress of patients through the Kent and Canterbury trial*

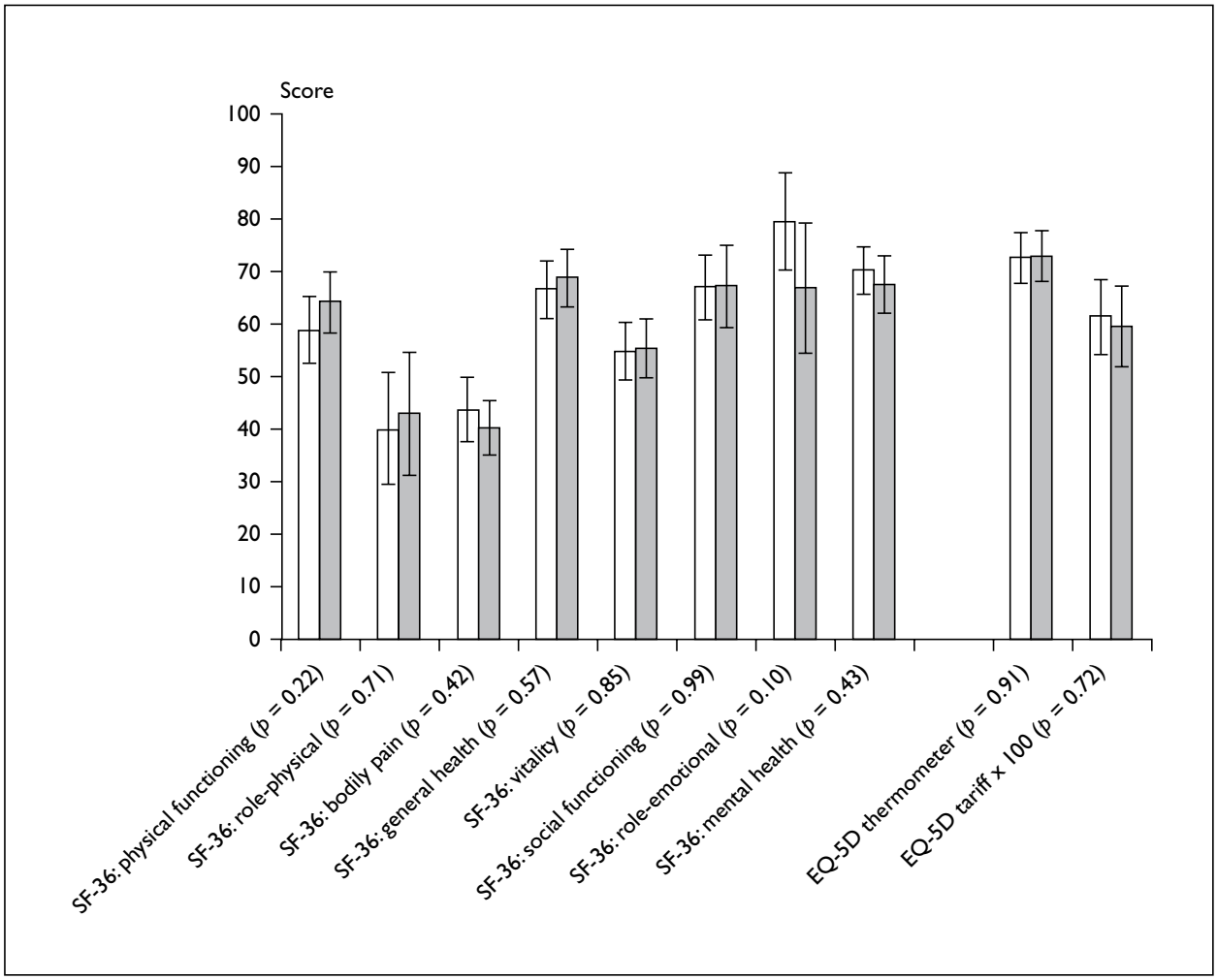


FIGURE 3 *Baseline SF-36 and EQ-5D data (mean scores, 95% CIs) (□, MRI group; ▩, no-MRI group)*

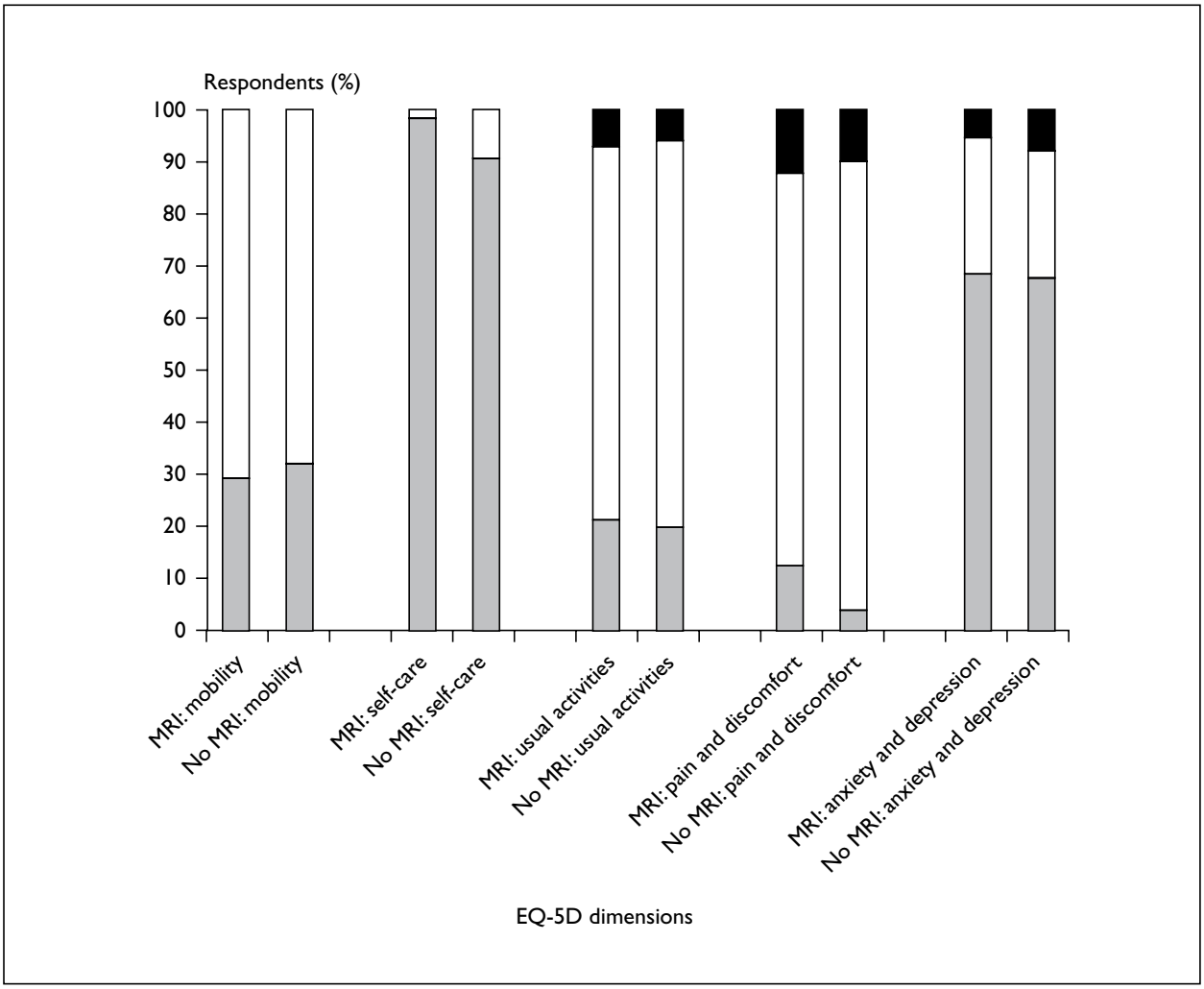


FIGURE 4 *Baseline EQ-5D data (■, Extreme problems; □, some problems; ▩, no problems)*

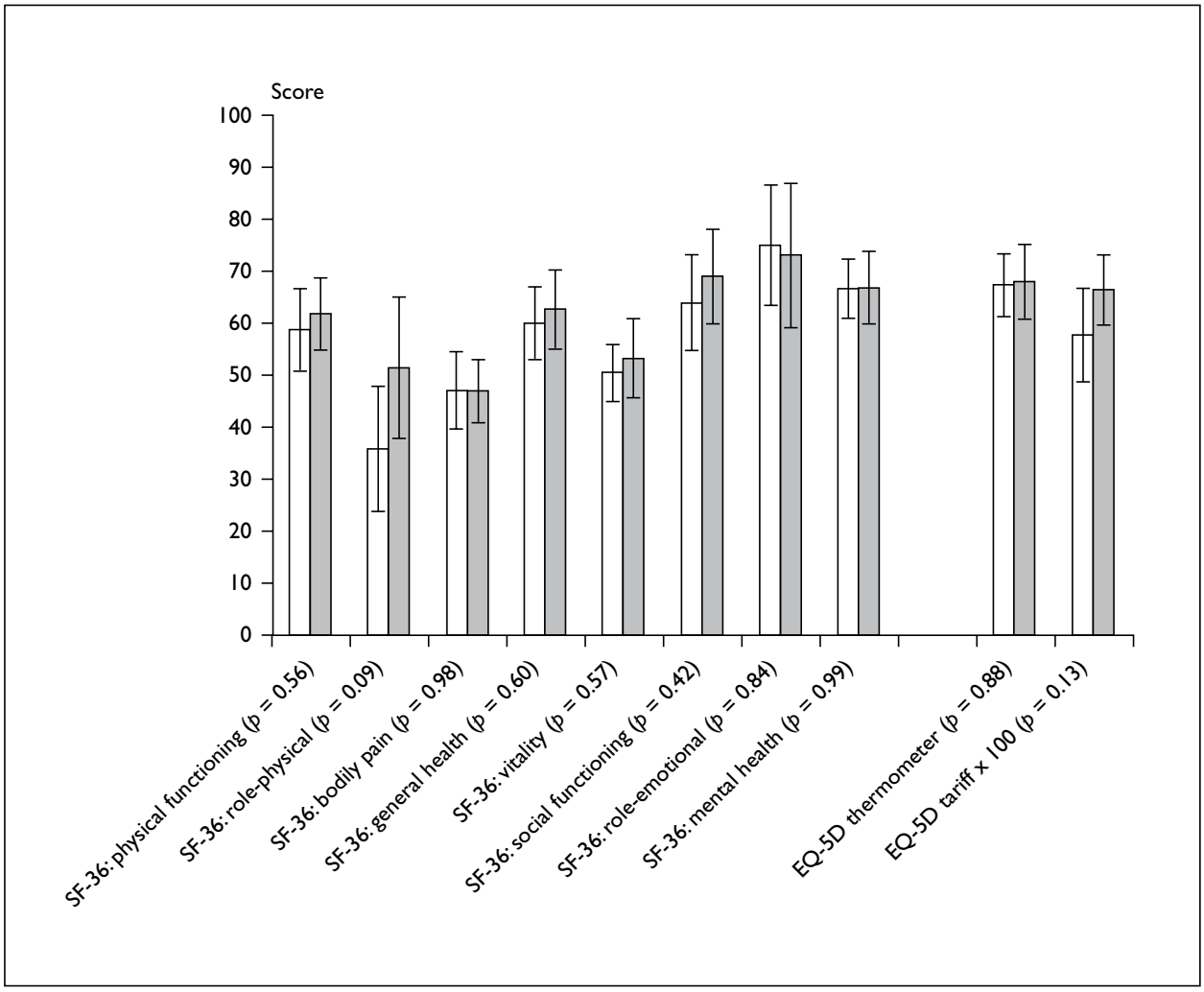


FIGURE 5 *SF-36 and EQ-5D data at 6 months (mean scores, 95% CIs) (□, MRI group; ▩, no-MRI group)*

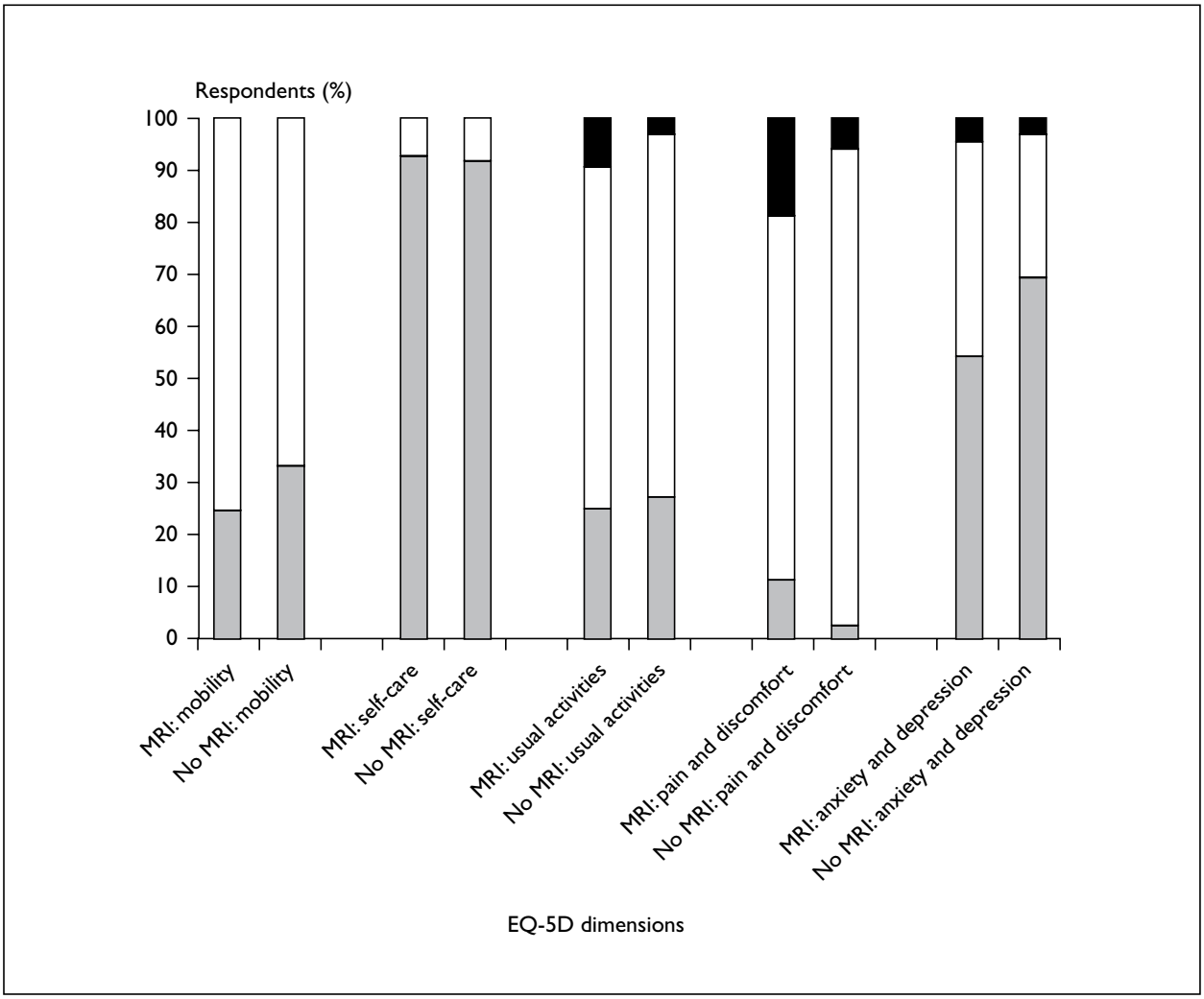


FIGURE 6 *EQ-5D data at 6 months (■, Extreme problems; □, some problems; ▨, no problems)*

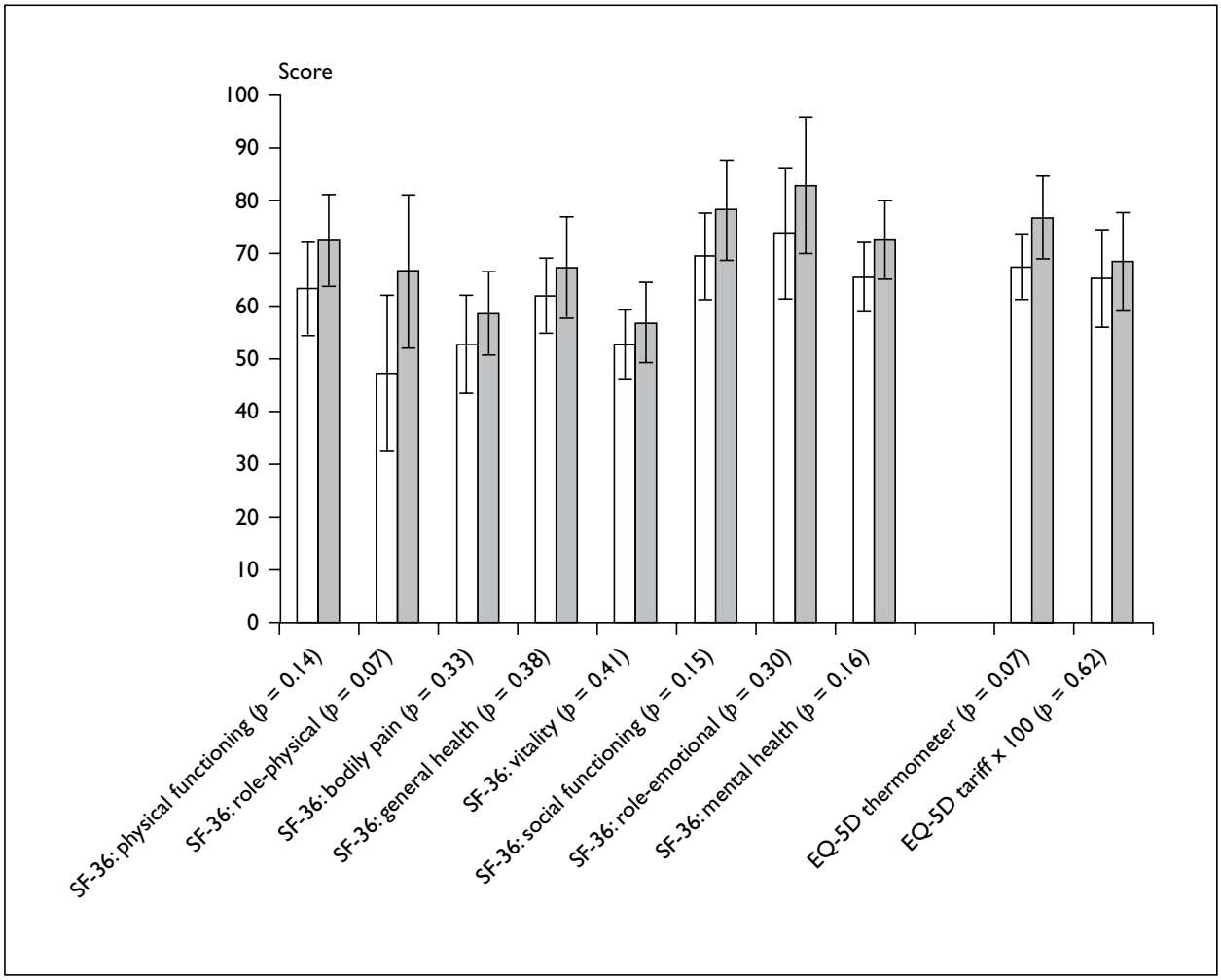


FIGURE 7 *SF-36 and EQ-5D data at 12 months (mean scores, 95% CIs) (□, MRI group; ▨, no-MRI group)*

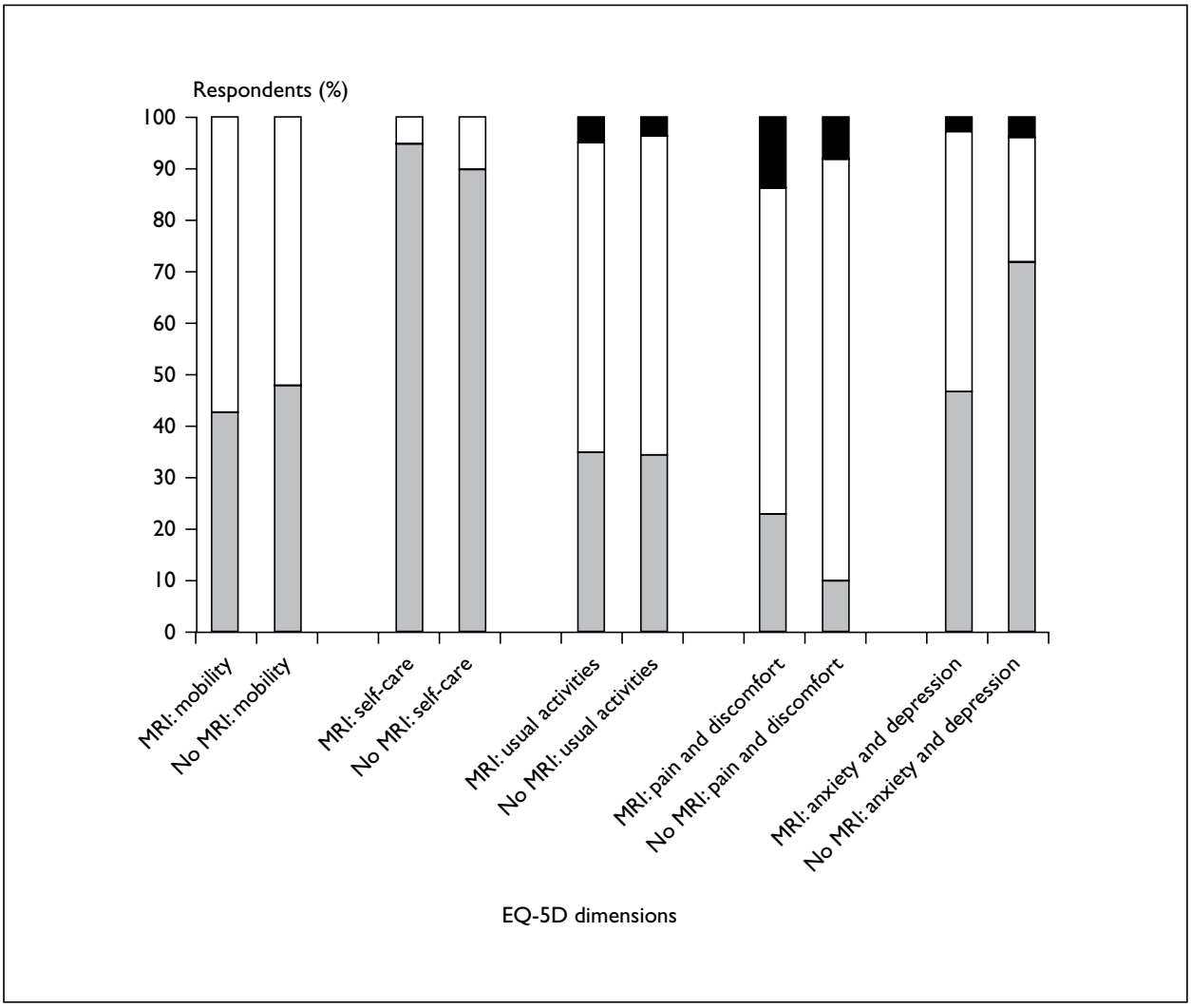


FIGURE 8 *EQ-5D data at 12 months (■, Extreme problems; □, some problems; ▩, no problems)*

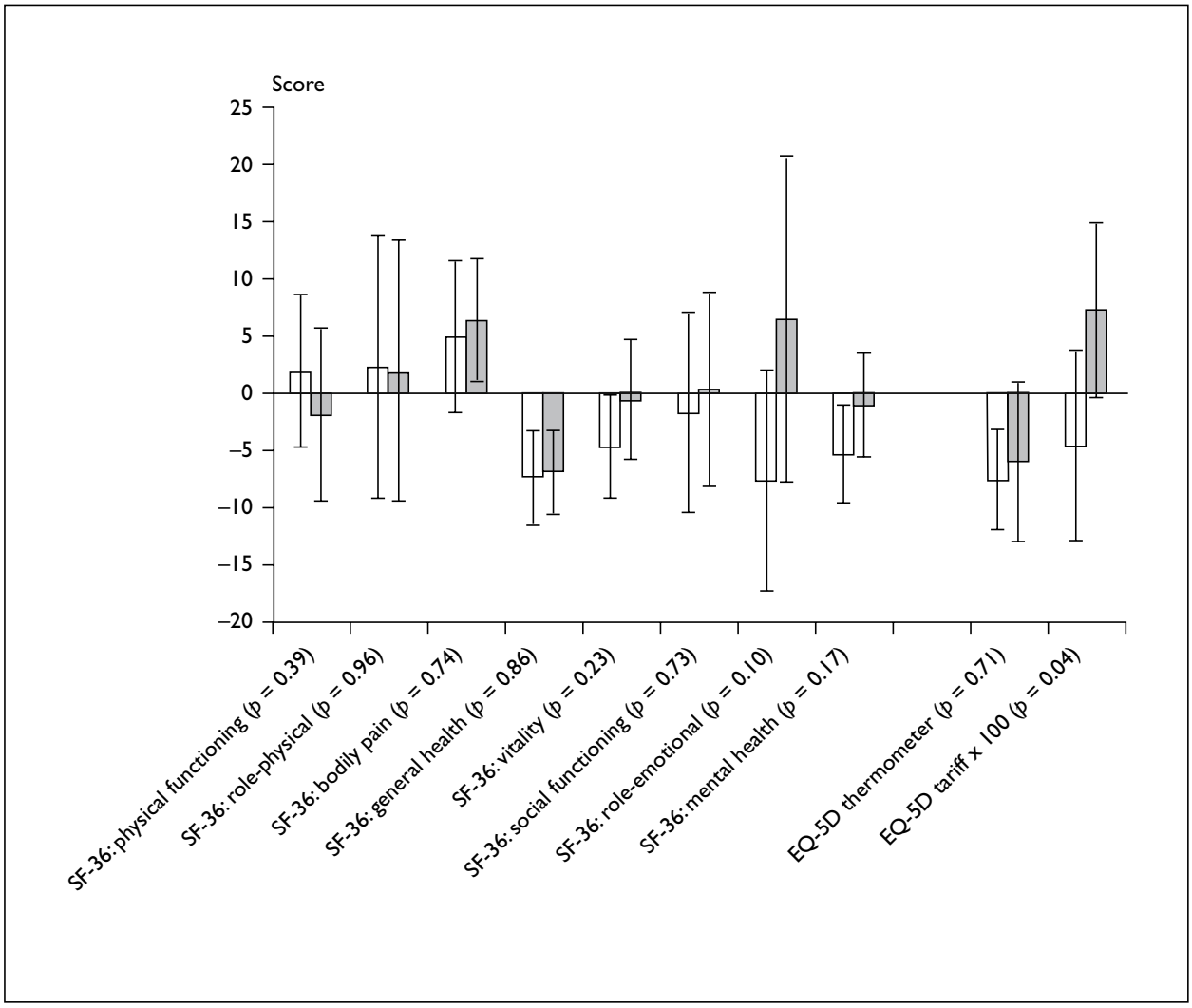


FIGURE 9 *Change from 0 to 6 months in SF-36 and EQ-5D data (mean scores, 95% CIs) (□, MRI group; ▩, no-MRI group)*

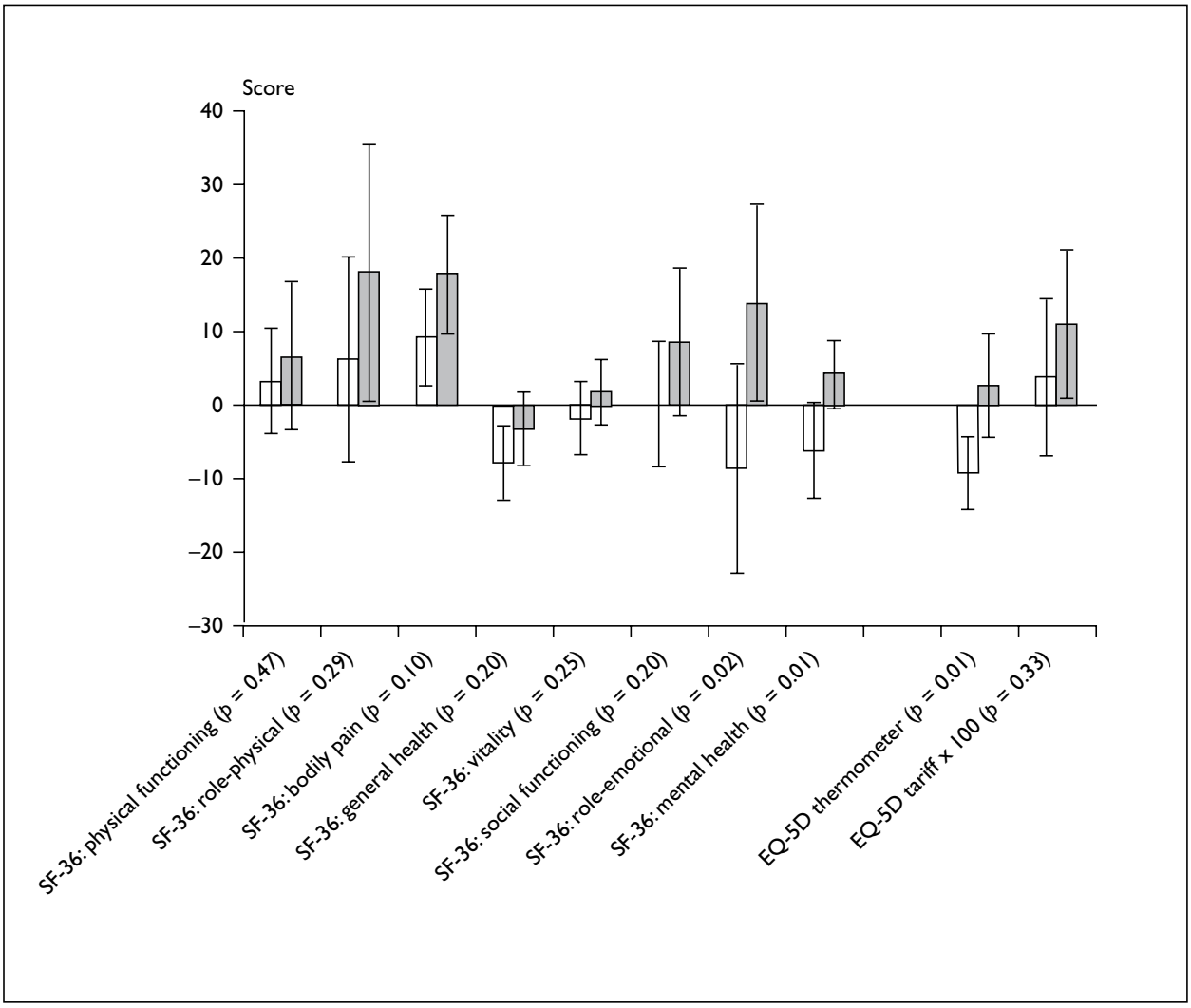


FIGURE 10 *Change from 0 to 12 months in SF-36 and EQ-5D data (mean scores, 95% CIs) (□, MRI group; ▩, no-MRI group)*

TABLE 2 *Estimation results from limited dependent variable model with sample selection: EQ-5D tariff*

	Main model			Sample selection mechanism		
	Coefficient estimate	**SE**	**Estimate/SE**	**Coefficient estimate**	**SE**	**Estimate/SE**
Baseline survey						
Sex	–0.0439	0.0659	–0.6662			
Trial arm	0.0389	0.0480	0.8104			
Source of referral	–0.0545	0.1056	–0.5161			
Age	–0.0065	0.0015	–4.3333[a]			
Constant	0.8541	0.2001	4.2684[a]			
6-month survey						
Sex	0.0435	0.0668	0.6512	–0.1247	0.1906	–0.6543
Trial arm	–0.0865	0.0666	–1.2988	0.4498	0.1732	2.5970[a]
Source of referral	0.0647	0.1580	0.4095			
Age	–0.0070	0.0033	–2.1212	0.0259	0.0104	2.4904
Constant	0.9010	0.1450	6.2138[a]	–0.2137	0.4901	–0.4360
12-month survey						
Sex	0.0032	0.0951	0.0337	–0.1130	0.1845	–0.6125
Trial arm	–0.0121	0.8001	–0.0151	0.4395	0.1685	2.6083[a]
Source of referral	–0.0323	0.1561	–0.2069			
Age	–0.0103	0.0039	–2.6410[a]	0.0221	0.0102	2.1667
Constant	1.0052	0.1865	5.3898[a]	–0.7174	0.4888	–1.4677
Random effects	–0.1120	0.6654	–0.1683	–0.3214	0.5512	–0.5831
Correlation (θ)				0.0806	0.4423	0.1822
Sigma	0.2663	0.0192	13.8698[a]			

SE, standard error

Maximised value of the log-likelihood: –1497.86

Number of observations: 104

[a] *Significant below 1% level in two-tailed test*

TABLE 3 *Estimation results from limited dependent variable model with sample selection: EQ-5D VAS score*

	Main model			Sample selection mechanism		
	Coefficient estimate	**SE**	**Estimate/SE**	**Coefficient estimate**	**SE**	**Estimate/SE**
Baseline survey						
Sex	–1.4642	3.8056	–0.3847			
Trial arm	–0.0382	3.2150	–0.0119			
Source of referral	–3.7891	5.5210	–0.6863			
Age	–0.2156	0.2614	–0.8248			
Constant	79.8895	7.8323	10.2000[a]			
6-month survey						
Sex	5.4404	5.0121	1.0855	–0.2061	0.2067	–0.9971
Trial arm	–1.3350	3.6870	–0.3621	0.3509	0.1879	1.8675
Source of referral	9.6573	8.5291	1.1323			
Age	–0.1125	0.2001	–0.5622	0.0221	0.0113	1.9558
Constant	70.3150	10.2256	6.8764[a]	–0.0484	0.5145	–0.0941
12-month survey						
Sex	1.5568	5.8132	0.2678	–0.1720	0.1961	–0.8771
Trial arm	–9.8754	4.8951	–2.0174	0.2764	0.1799	1.5364
Source of referral	7.5778	8.9968	0.8423	0.0194		
Age	–0.2564	0.2669	–0.9607	0.0194	0.0109	1.7798
Constant	85.5594	10.2356	8.3590[a]	–0.6382	0.5101	–1.2511
Random effects	–0.2193	0.7785	–0.2817	1.001	0.9896	1.0115
Correlation (θ)				0.1825	0.4897	0.3727
Sigma	16.7096	1.1642	14.3529[a]			

Maximised value of the log-likelihood: –1501.32

Number of observations: 103

[a] *Significant below 1% level in two-tailed test*

TABLE 4 *Estimation results from limited dependent variable model with sample selection: SF-36 physical functioning score*

	Main model			Sample selection mechanism		
	Coefficient estimate	**SE**	**Estimate/SE**	**Coefficient estimate**	**SE**	**Estimate/SE**
Baseline survey						
Sex	0.5012	5.6231	0.0891			
Trial arm	–4.4561	4.4473	–1.0020			
Source of referral	–2.9632	8.0123	–0.3698			
Age	–0.5154	0.2177	–2.3675			
Constant	82.0300	8.4465	9.7117[a]			
6-month survey						
Sex	13.7695	5.9587	2.3108	–0.2332	0.2014	–1.1579
Trial arm	–2.5657	4.9865	–0.5145	0.3645	0.1823	1.9995
Source of referral	–8.3363	10.2541	–0.8130			
Age	–0.7991	0.3001	–2.6628[a]	0.0240	0.0110	2.1818
Constant	79.9900	10.0021	7.9973[a]	–0.1407	0.5068	–0.2776
12-month survey						
Sex	7.0021	6.5859	1.0632	–0.2236	0.1923	–1.1628
Trial arm	–8.0051	6.4598	–1.2392	0.2850	0.1758	1.6212
Source of referral	–13.1130	11.9778	–1.0948			
Age	–0.7214	0.3188	–2.2629	0.0239	0.0106	2.2547
Constant	97.2251	12.9590	7.5025[a]	–0.7466	0.5056	–1.4767
Random effects	0.0010	0.0080	0.1250	–0.0315	0.5512	–0.0571
Correlation (θ)				0.1200	0.1514	0.7926
Sigma	22.3220	1.5774	14.1511[a]			

Maximised value of the log-likelihood: –1488.26

Number of observations: 108

[a] *Significant below 1% level in two-tailed test*

TABLE 5 *Estimation results from limited dependent variable model with sample selection: SF-36 bodily pain score*

	Main model			Sample selection mechanism		
	Coefficient estimate	**SE**	**Estimate/SE**	**Coefficient estimate**	**SE**	**Estimate/SE**
Baseline survey						
Sex	–1.4351	4.1142	–0.3488			
Trial arm	4.1562	3.9523	1.0516			
Source of referral	–5.5969	7.5582	–0.7405			
Age	–0.5254	0.2046	–2.5679[a]			
Constant	60.0125	7.9980	7.5034[a]			
6-month survey						
Sex	7.0275	5.3320	1.3180	–0.2801	0.2024	–1.3839
Trial arm	1.4961	4.9968	0.2994	0.3331	0.1839	1.8113
Source of referral	1.1323	11.0010	0.1029			
Age	–0.6631	0.2500	–2.6524[a]	0.0271	0.0110	2.4636
Constant	65.5550	11.3112	5.7956[a]	–0.2742	0.5142	–0.5333
12-month survey						
Sex	–5.7362	8.0115	–0.7160	–0.2179	0.1923	–1.1331
Trial arm	–6.3312	7.2564	–0.8725	0.2951	0.1769	1.6682
Source of referral	14.4493	14.3998	1.0034			
Age	–0.9011	0.3594	–2.5072	0.0220	0.0106	2.0755
Constant	92.8820	15.7876	5.8832[a]	–0.8012	0.5152	–1.5551
Random effects	–0.1130	0.2130	–0.5305	–0.0110	0.0252	–0.4365
Correlation (θ)				0.0081	0.0911	0.0889
Sigma	20.6041	1.4610	14.1027[a]			

Maximised value of the log-likelihood: –1514.66

Number of observations: 106

[a] *Significant below 1% level in two-tailed test*

TABLE 6 *Estimation results from limited dependent variable model with sample selection: SF-36 general health score*

	Main model			Sample selection mechanism		
	Coefficient estimate	**SE**	**Estimate/SE**	**Coefficient estimate**	**SE**	**Estimate/SE**
Baseline survey						
Sex	1.2048	4.1201	0.2924			
Trial arm	–1.6821	3.8638	–0.4353			
Source of referral	–4.7891	7.0254	–0.6817			
Age	–0.1491	0.1855	–0.8038			
Constant	73.7359	7.7089	9.5650[a]			
6-month survey						
Sex	1.3134	5.5089	0.2384	–0.2625	0.2024	–1.2969
Trial arm	–2.4249	5.1106	–0.4745	0.3287	0.1824	1.8021
Source of referral	10.9558	10.6263	1.0310			
Age	–0.0748	0.2661	–0.2811	0.0271	0.0110	2.4636
Constant	63.9379	10.8458	5.8952[a]	–0.1713	0.5068	–0.3380
12-month survey						
Sex	2.7817	6.5627	0.4239	–0.2510	0.1922	–1.3059
Trial arm	–4.8997	5.9875	–0.8183	0.3208	0.1757	1.8258
Source of referral	6.8458	11.3362	0.6039			
Age	–0.5646	0.3044	–1.8548	0.0228	0.0106	2.1509
Constant	86.4691	12.5397	6.8956[a]	–0.7380	0.5054	–1.4602
Random effects	–0.0990	0.1020	–0.9706	0.2251	0.4632	0.4860
Correlation (θ)				0.1011	0.0991	1.0202
Sigma	19.9023	1.3837	14.3834[a]			

Maximised value of the log-likelihood: –1500.12

Number of observations: 108

[a] *Significant below 1% level in two-tailed test*

TABLE 7 *Estimation results from limited dependent variable model with sample selection: SF-36 vitality score*

	Main model			Sample selection mechanism		
	Coefficient estimate	**SE**	**Estimate/SE**	**Coefficient estimate**	**SE**	**Estimate/SE**
Baseline survey						
Sex	2.5799	4.0150	0.6426			
Trial arm	0.0065	3.7640	0.0017			
Source of referral	–4.9767	6.8485	–0.7267			
Age	–0.5948	0.1808	–3.2898[a]			
Constant	75.0965	7.5133	9.9951[a]			
6-month survey						
Sex	1.5578	4.6463	0.3353	–0.2717	0.2022	–1.3437
Trial arm	–2.0817	4.3083	–0.4832	0.3413	0.1819	1.8763
Source of referral	–7.6922	8.9674	–0.8578			
Age	–0.5490	0.2245	–2.4454	0.0269	0.0110	2.4455
Constant	72.5730	9.1512	7.9304[a]	–0.1698	0.5068	–0.3350
12-month survey						
Sex	7.7168	5.1380	1.5019	–0.2395	0.1919	–1.2480
Trial arm	–2.3044	4.6869	–0.4917	0.3053	0.1751	1.7436
Source of referral	6.7570	8.9310	0.7566			
Age	–0.5064	0.2374	–2.1331	0.0232	0.0106	2.1887
Constant	68.7798	9.6708	7.1121[a]	–0.7380	0.5051	–1.4611
Random effects	–0.1020	0.9901	–0.1030	0.3051	0.4891	0.6238
Correlation (θ)				0.1131	0.1257	0.8998
Sigma	19.4054	1.3404	14.4773[a]			

Maximised value of the log-likelihood: –1499.07

Number of observations: 108

[a] *Significant below 1% level in two-tailed test*

TABLE 8 *Estimation results from limited dependent variable model with sample selection: SF-36 social functioning score*

	Main model			Sample selection mechanism		
	Coefficient estimate	**SE**	**Estimate/SE**	**Coefficient estimate**	**SE**	**Estimate/SE**
Baseline survey						
Sex	1.5362	6.2325	0.2465			
Trial arm	–0.9880	5.8820	–0.1680			
Source of referral	–4.9103	10.5764	–0.4643			
Age	–0.1629	0.2844	–0.5728			
Constant	75.5822	11.7831	6.4145[a]			
6-month survey						
Sex	1.4522	9.3362	0.1555	–0.2642	0.2023	–1.3060
Trial arm	–3.8568	8.6699	–0.4448	0.3271	0.1824	1.7933
Source of referral	–3.1529	17.8943	–0.1762			
Age	–0.3042	0.4531	–0.6714	0.0254	0.0111	2.2883
Constant	84.2876	18.4250	4.5746[a]	–0.1219	0.5089	–0.2395
12-month survey						
Sex	–1.8910	9.9326	–0.1904	–0.2337	0.1920	–1.2172
Trial arm	–11.1615	9.1733	–1.2167	0.2947	0.1755	1.6792
Source of referral	17.2446	17.7420	0.9720			
Age	–0.2038	0.4626	–0.4406	0.0221	0.0106	2.0849
Constant	92.2879	18.9807	4.8622[a]	–0.7004	0.5065	–1.3828
Random effects	–0.0834	1.0002	–0.0834	–0.0135	0.9980	–0.0135
Correlation (θ)				0.2156	0.1895	1.1377
Sigma	29.6333	2.3436	12.6444[a]			

Maximised value of the log-likelihood: –1501.55

Number of observations: 107

[a] *Significant below 1% level in two-tailed test*

TABLE 9 *Estimation results from limited dependent variable model with sample selection: SF-36 mental health score*

	Main model			Sample selection mechanism		
	Coefficient estimate	**SE**	**Estimate/SE**	**Coefficient estimate**	**SE**	**Estimate/SE**
Baseline survey						
Sex	2.4732	3.7549	0.6587			
Trial arm	3.1184	3.5196	0.8860			
Source of referral	–3.9087	6.4063	–0.6101			
Age	–0.1058	0.1691	–0.6257			
Constant	70.0881	7.0268	9.9744[a]			
6-month survey						
Sex	–0.2893	4.6418	–0.0623	–0.2658	0.2022	–1.3145
Trial arm	0.0456	4.3009	0.0106	0.3547	0.1824	1.9446
Source of referral	7.2503	8.9443	0.8106			
Age	–0.3948	0.2241	–1.7617	0.0261	0.0110	2.3727
Constant	81.1301	9.1517	8.8650[a]	–0.1481	0.5073	–0.2919
12-month survey						
Sex	5.0897	5.1703	0.9844	–0.2462	0.1920	–1.2823
Trial arm	–0.8285	4.7653	–1.2231	0.2911	0.1756	1.6577
Source of referral	10.0647	9.1197	1.1036			
Age	–0.4351	0.2424	–1.7950	0.0240	0.0106	2.2642
Constant	83.8182	9.8481	8.5111[a]	–0.7634	0.5058	–1.5093
Random effects	0.0020	0.0056	0.3571	–0.0256	0.1200	–0.2133
Correlation (θ)	0.1599	0.0998	1.6022			
Sigma	18.1528	1.2438	14.5946[a]			

Maximised value of the log-likelihood: –1498.73

Number of observations: 107

[a] *Significant below 1% level in two-tailed test*

TABLE 10 *NHS resource use and costs (£) over 12 months (base case)*

	MRI (*n* = 59)	**MRI 95% CI**	**No-MRI (*n* = 59)**	**No-MRI 95% CI**	**Difference between arms**	**95% CI**
Proportion undergoing surgery	0.41	0.28 to 0.54	0.71	0.58 to 0.82	–0.30	–0.48 to –0.14
Mean (SD) number of outpatient visits	2.61 (1.34)	2.26 to 2.96	2.29 (1.27)	1.96 to 2.62	0.32	–0.16 to 0.79
Mean (SD) number of GP visits	2.25 (2.21)	1.67 to 2.83	1.62 (1.88)	1.13 to 2.11	0.63	–0.12 to 1.38
Mean (SD) number of physiotherapy sessions	4.36 (6.11)	2.77 to 5.95	3.44 (4.96)	2.15 to 4.73	0.92	–1.11 to 2.95
Mean (SD) total NHS costs[a]	756 (809)	609 to 1121	708 (607)	594 to 926	48.12	–181.41 to 335.49
Mean (SD) total patient costs[a]	141 (100)	117 to 170	137 (90)	114 to 161	3.76	–28.93 to 40.94
Mean (SD) total NHS + patient costs[a]	897 (886)	730 to 1227	845 (678)	707 to 1077	51.88	–197.53 to 369.98

[a] *Bootstrap comparison of means, 95% CI (bias corrected and accelerated method, 2000 replications)*

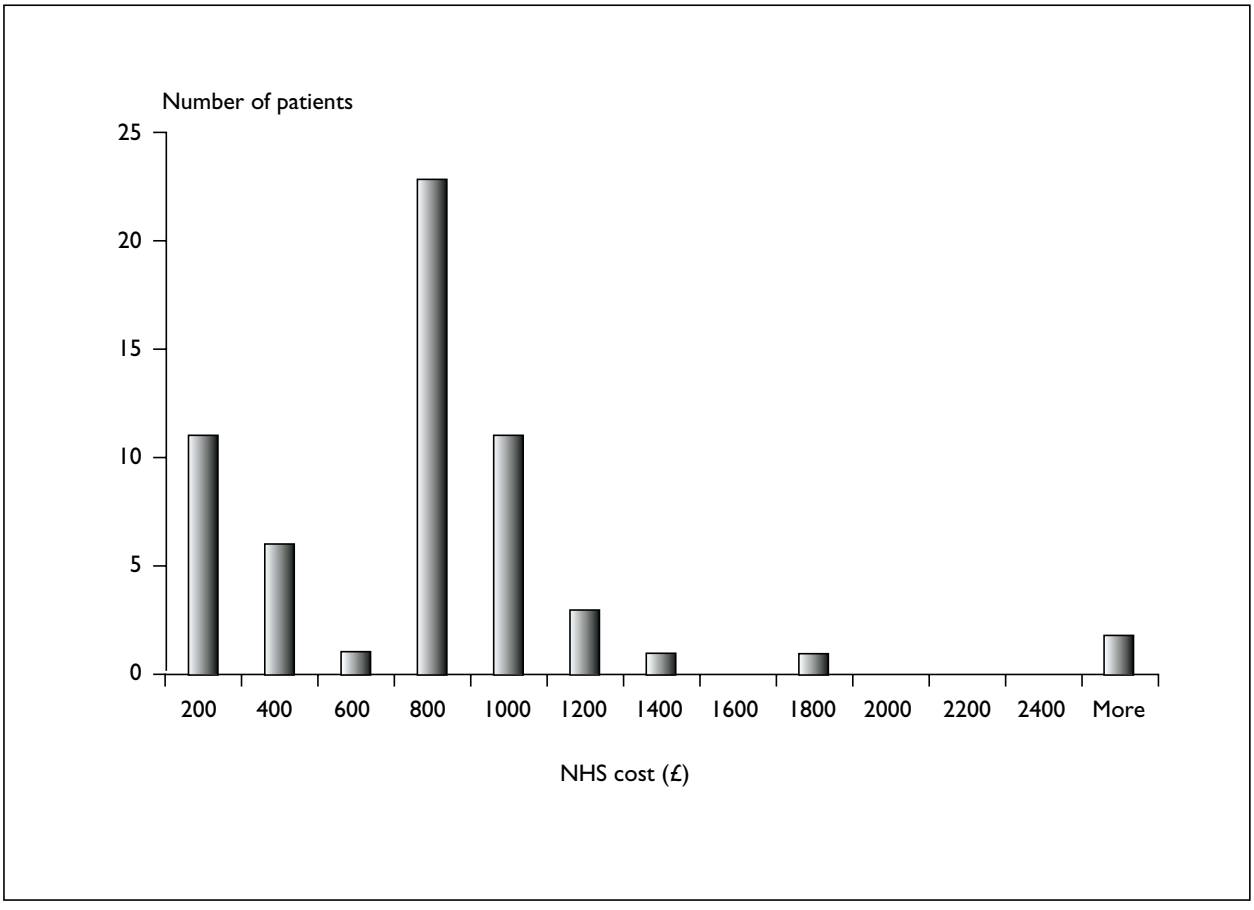


FIGURE 11 *Frequency distribution of NHS costs: MRI arm*

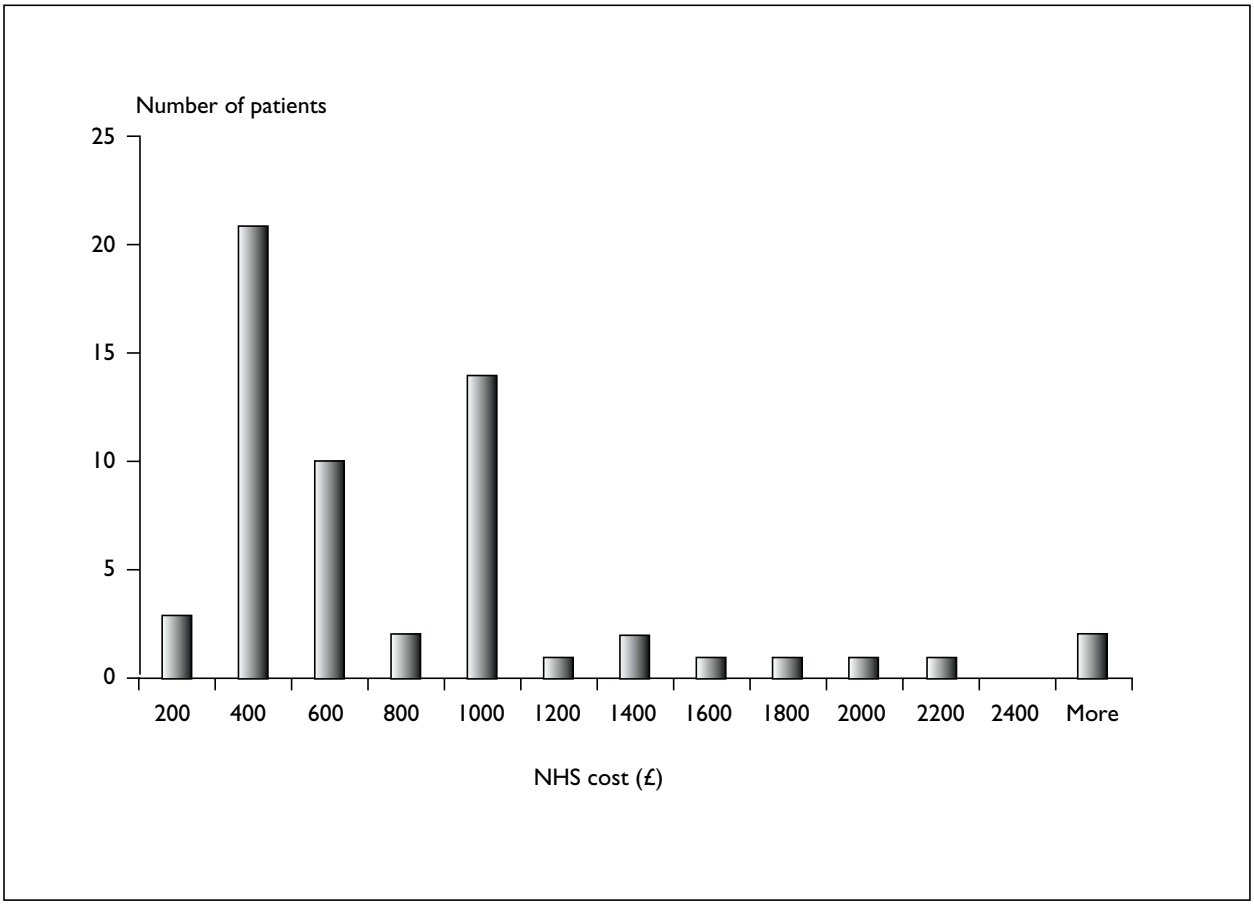


FIGURE 12 *Frequency distribution of NHS costs: no-MRI arm*

Chapter 5

The diagnostic accuracy of clinical investigation and MRI in acute knee injuries

Introduction

The research reported in chapter 4 compared alternative diagnostic strategies in patients presenting with knee problems who were being considered for arthroscopy by the orthopaedic surgeon. However, it is important to also consider the initial clinical investigation patients receive when first presenting with a knee injury. If routine clinical investigation were itself associated with a high level of diagnostic accuracy, then the need for either diagnostic arthroscopy or MRI could be questioned in some patients. This is the focus for the substudy reported in this chapter, which addressed the following research question:

> What is the diagnostic accuracy of the initial clinical investigation of knee injuries (by orthopaedic surgeons without MRI and radiologists with access to MRI) and to what extent does this vary by grade and experience of surgeon?

The data reported in this chapter were collected as part of an additional, separate clinical study conducted at St Thomas' Hospital, London, UK in which MRI scans were undertaken in a cohort of patients with acute knee injuries. The objective of this study was to consider how the diagnoses of patients presenting with knee injuries varied between an orthopaedic trainee (with access to the patient but not the MRI scan), a consultant knee specialist (with access to the patient but not the MRI scan) and a consultant radiologist (with access to the MRI scan but not the patient). In addition, the study aimed to measure the diagnostic accuracy of each of these clinicians by comparing their diagnosis with a 'final' or 'reference standard' diagnosis.

Methods

Patient recruitment into the study took place through a specialist knee clinic. The clinic received patients with acute knee injuries referred from the Accident and Emergency Department. Following receipt of Ethics Committee approval, all patients who attended the knee clinic between April 1996 and July 1997 were assessed for suitability for inclusion in the study. The aim was for the patient cohort to be representative of the range of knee injuries seen in an Accident and Emergency Department.

The eligibility criteria were similar to those of the Kent and Canterbury trial; patients were defined as suitable if:

- there had been no previous major surgery in the injured knee, such as knee replacement (previous arthroscopy and partial menisectomy did not exclude patients from the trial)
- there was no pre-existing chronic knee pathology
- there was no serious condition requiring immediate attention, for example, a serious knee infection
- there was no history or current experience of recurrent locking of the knee
- they were aged between 16 and 55 years.

Written consent was sought at the initial clinic visit. All study patients were referred for an MRI scan, which was usually performed within 48 hours (median wait = 2 days), and a consultant radiologist reported on each patient ('radiologist diagnosis'). Neither the MRI scan nor the radiologist's report (either in paper form or through the hospital computer system) was available to the orthopaedic clinicians during the primary data collection phase of the study. The consultant knee specialist reviewed each patient 1 week after the MRI scan and recorded a provisional diagnosis ('knee specialist diagnosis'). During the same hospital visit, patients were also seen separately by one of two post-final fellowship senior registrars who independently provided a diagnosis ('senior registrar diagnosis'). Once the three diagnoses had been obtained, primary data collection was complete and patient management and follow-up continued until the patients' knee injuries had recovered.

At the end of data collection, two clinical members of the research team (NCH, FWH) determined a final reference standard diagnosis for all study patients. The process involved the clinical review of three separate sets of information for every study patient:

- the patients hospital notes, providing data on the patient's medical history, symptoms at presentation, follow-up information, subsequent clinical progress, results of surgery (if relevant), etc.
- the MRI scans taken following initial presentation and any other images relating to additional radiographic procedures undertaken
- the radiologist's interpretation of the MRI scans (always provided by a consultant radiologist with experience in reading MRI images) and radiology reports for any other examinations.

A final diagnosis was reached for each patient by taking all of this information into account. This represented the reference standard diagnosis used in assessing diagnostic accuracy, and a simple comparison was made of the diagnosis recorded by each clinician and the final reference standard diagnosis.

The sample for this analysis comprised 114 patients who received an MRI scan and for whom a final reference standard diagnosis was available. The mean age of the sample was 28 years (SD = 6.17, range 16 to 47) and 82 (72%) were male. A radiology report, provided by the same consultant radiologist, was available for all 114 patients, and the orthopaedic trainees also saw all patients between them. However, the knee specialist was only able to see 71 of the 114 patients.

Results

Figure 13 illustrates the spectrum of diagnoses made by the orthopaedic trainees, the consultant knee specialist and the radiologist in the 71 knees seen in clinic by the knee specialist. The orthopaedic trainees diagnosed meniscus damage in 44% of the injured knees and were less likely to diagnose an injury to the ACL. The knee specialist, however, was more likely to diagnose an injury to the ACL. He was also more likely to say that the diagnosis was unclear at the initial assessment and would require repeat examinations. The consultant radiologist diagnosed a meniscal lesion extending to the articular surface in 33% of the knees, but less than two-thirds of these had clinical symptoms. This warranted meniscectomy in those without symptoms during the study period with subsequent follow-up. The MRI scan had low sensitivity for injuries involving the patello-femoral mechanism.

The original diagnoses of the clinicians and consultant radiologist were compared to the final reference diagnoses (*Table 11*). The orthopaedic trainees obtained a correct diagnosis in

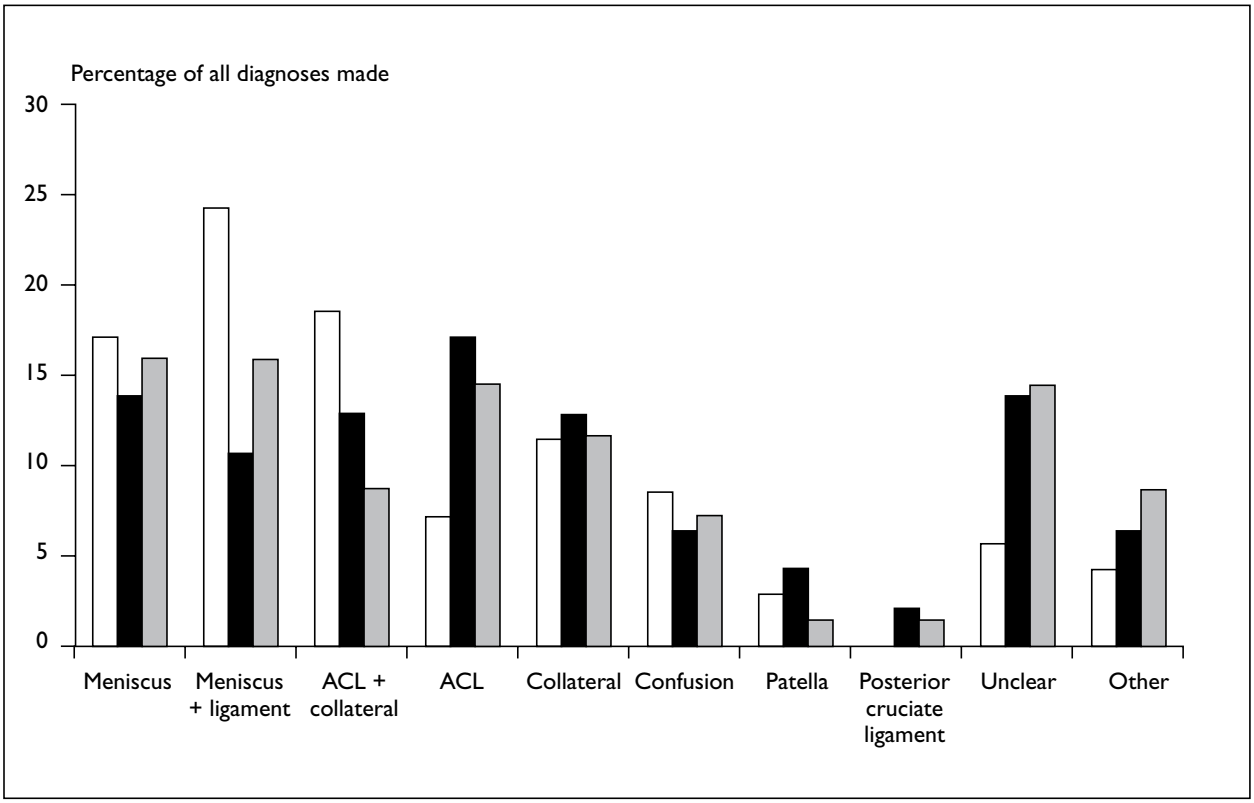


FIGURE 13 *Diagnoses of the orthopaedic trainees (□), consultant knee specialist (■) and consultant radiologist (▩)*

TABLE 11 *Diagnostic accuracy*

	'Correct' diagnosis	'Incorrect' diagnosis	'Incomplete' diagnosis	Total
Senior registrar	50 (44%)	39 (34%)	25 (22%)	114
Knee specialist	51 (72%)	6 (8%)	14 (20%)	71
Consultant radiologist	77 (68%)	6 (5%)	31 (27%)	114

only 44% of the patients overall, although this was noted to improve with experience. For the first 20 cases, the correct diagnosis was obtained in only 25% of the knees, which contrasts with the final 20 cases in which the trainees made the correct and complete diagnosis in 70% of the patients and only missed one important lesion (a lateral meniscal tear). The knee specialist recorded a correct and complete diagnosis in 72% of the patients he saw and only misdiagnosed six (8%) patients, five of which were false-negatives. The consultant radiologist correctly diagnosed 68% of the patients. In 27% of cases, his diagnosis was incomplete or misleading, but in only six patients (5%) was an important element of the diagnosis missed. *Table 12* provides some examples of missed diagnoses by both the knee specialist and the consultant radiologist.

TABLE 12 *Incorrect diagnoses by knee specialist and consultant radiologist*

Knee specialist	Consultant radiologist
Missed lateral meniscal tear	Missed partial ACL tear
Missed lateral meniscal tear	Missed partial posterior cruciate ligament tear
Missed tibial plateau bone cyst and bruise	Missed synovitis
Missed undisplaced lateral tibial plateau fracture	Missed medial meniscal tear
Missed partial ACL tear	Missed patella chondral injury
Over-diagnosis of ACL tear	Missed patella dislocation and chondral injury

Discussion

The data from this study suggests that, when compared to non-specialists or to radiologists reporting on an MRI scan, diagnostic accuracy is higher for knee specialists following clinical examination of the knee. In broad terms, these results are comparable to previous studies.[61,62] It has been demonstrated previously that an accurate history and examination carried out by an experienced knee specialist leads to the incorrect diagnosis and management of patients in only a few cases.[41,63–66] The decision of whether to perform arthroscopic surgery of the knee should, therefore, be based on the patient's history, physical examination and radiographs. Furthermore, the final decision regarding surgery in doubtful cases should be made after the patient has been reviewed and reassessed on several occasions, and MRI should perhaps be reserved for patients who either have an unclear diagnosis or have failed to improve with the previous management plan. Reliance on the result of an MRI scan without the necessary skilled clinical assessment may lead to the mismanagement of patients.[15,67,68]

One of the weaknesses of this substudy concerns the generalisability of the findings. Data were generated using a single radiologist, a single knee specialist and two orthopaedic trainees. In order to explore the extent to which the results of this study were specific to the St Thomas' Hospital radiologist involved and, more generally, to consider issues of variation in radiologists' interpretations of MRI scans, additional research was conducted. This research explored the level of agreement between radiologists when interpreting knee MRI images, and is reported in the next chapter.

Chapter 6

An investigation of variation between radiologists in knee MRI interpretation

Introduction

In assessing the quality of radiographic imaging examinations, the two distinct components are the acquisition of satisfactory images and the interpretation of the images by a radiologist. It has been shown that when more than one viewer interprets the same images there may be variations in the interpretation. This is known as 'inter-viewer' variation.[69–71] In addition, if the same viewer interprets the same images on different occasions there may again be variations in the interpretation. This is known as 'intra-viewer' variation. If different images are used to demonstrate the same area of examination there may be differences in interpretation relating to 'inter-image' variations. As indicated in chapter 3, it has been shown that the specificity and sensitivity of MRI is dependent on the reporting radiologist and the strength of the magnetic field.[14]

The previous chapter discussed the accuracy of the diagnoses made by both the orthopaedic surgeons and the radiologist at St Thomas' Hospital compared with the final orthopaedic reference diagnosis. The final diagnosis was made by two of the authors of this report (NCH, FWH) who had access to each patient's hospital notes, the report of MRI examinations and details of the clinical follow-up. The aim of the substudy reported in this chapter was to determine whether the diagnosis of the images would differ significantly if consultant radiologists in a DGH setting interpreted the MRI examinations from St Thomas' Hospital.

Methods

Seven consultant radiologists who all worked in the MRI units at DGHs independently interpreted MRI images from a sample of study patients recruited at St Thomas' Hospital. A data collection pro-forma for the scoring of images was designed in consultation with both orthopaedic and radiology colleagues. This eliminated the need for free text responses and allowed the data to be easily entered into a database for analysis. The aim was to have a very simple form that would be quick to complete. Each viewer completed a separate form for each examination (see appendix 5 for the pro-forma used).

Pilot study

A small pilot study was undertaken in order to determine:

- the sample size that was required for the main study
- the length of time required by radiologists to undertake such reports (in order to determine the additional cost of the study and the time commitment for radiologists)
- the suitability of the pro-forma.

The original MRI films for ten patients were scored separately by the consultant radiologist in the specialist centre and a consultant radiologist in a DGH who regularly made reports of orthopaedic MRI scans, but was not already involved in the MRI trial.

Sample size calculation

Based on the results of the pilot study, a formal sample size calculation was undertaken using data for one area of the knee, lateral collateral ligaments. The study was powered to detect a difference between the specialist and DGH centres of 0.178 in the proportion of responses indicating an abnormality (80% power, 0.05 significance level). The minimum sample required was 71 patients.[72] A pragmatic decision was taken to allow for up to 10% errors or omissions in the completion of forms and, thus, a sample of images of 80 patients was required for the main comparative study. Sample size calculations that were undertaken for other areas of the knee gave broad support for a sample of this size.

Main study

DGHs were identified which had an MRI unit installed that was in routine clinical use. A written request to take part in this study was made to the heads of nine of these MRI units, and positive responses were received from eight hospitals. Further details of the study, a copy of the scoring form, the payment that was available and the timescale that was required for the images to be

read were sent to these hospitals. The consultant radiologists in five of these hospitals agreed to participate. The hospitals were in Sunderland, Aberdeen, the Isle of Wight, Carshalton and Reading, and, from these DGHs, seven viewers (consultant radiologists) were recruited.

The MRI examinations of 80 patients who had been recruited to the study at St Thomas' Hospital were used for which the final orthopaedic diagnosis was known and the images and details of the information provided on the request form for the MRI examination were available. The images for the first 80 examinations that fulfilled these criteria were used. Each film envelope was labelled with the patient's trial number and this number was recorded on the scoring pro-forma. The request form details were included in the film envelopes. The 80 sets of hard-copy MRI images were sent by mail to each DGH in turn, and the radiologists were each asked to complete the study within 2 weeks and to send the films directly to the radiologist in the next DGH using address labels that were provided. Instructions for viewing and scoring the images were sent to all viewers (see appendix 6).

Finally, the orthopaedic registrar at the specialist centre transferred the final reference orthopaedic diagnoses to the format of the pro-forma used by the radiologists. For each area of the knee, the level of agreement (measured using weighted kappa) was assessed between the final orthopaedic diagnosis and the diagnosis of each of the eight radiologists (including the specialist centre radiologist). In addition, similar comparisons were made between the responses of the specialist centre radiologist and those of each of the seven radiologists in the DGHs. This allowed the mean (and SD) for the weighted kappa statistics to be calculated for each area of the knee. The weights used for the calculations are shown in *Table 13*. The categories suggested by Altman were used for the interpretation of the weighted kappa statistics:[57]

Poor agreement: kappa < 0.20
Fair agreement: kappa 0.21–0.40
Moderate agreement: kappa 0.41–0.60
Good agreement: kappa 0.61–0.80
Very good agreement: kappa 0.81–1.00

Results

The mean weighted kappa scores for the comparison between the final orthopaedic diagnosis and the diagnoses of all eight radiologists for each anatomical area of the knee are shown in *Figure 14*. A 'moderate' level of agreement was obtained for the body and posterior third of the medial meniscus and the ACL. The agreement was 'fair' for the anterior third of both the medial and lateral menisci, the posterior third of the cruciate ligament and the medial collateral ligament. However, agreement was 'poor' for the bony surfaces. Joint effusion was not assessed in the final orthopaedic diagnosis and no comparisons were, therefore, possible.

TABLE 13 *Weightings used for the calculation of the weighted kappa statistic*

Area of the knee	Weights
Medial meniscus and lateral meniscus	Not seen and normal = 1 Degenerative changes = 2 Tear = 3
Cruciate ligament	Not seen and normal = 1 Partial tear = 2 Complete rupture = 3
Collateral ligament	Not seen and normal = 1 Sprain = 2 Partial tear = 3 Complete tear = 4
Bony surfaces	Normal = 1 Degenerative changes = 2 Bruise = 3 Fracture = 4
Joint effusion	Not present = 1 Small = 2 Large = 3

The mean weighted kappa scores for the comparison between the responses of the specialist centre radiologist and those of the DGH radiologists for each anatomical area of the knee are shown in *Figure 15*. On average, there was 'moderate' or 'good' agreement for all areas except the lateral collateral ligament and the presence of joint effusion, which had 'poor' and 'fair' levels of agreement, respectively. For all areas of the knee, except the lateral collateral ligament, the kappa scores were higher for the comparison between the specialist centre radiologist and the seven DGH radiologists than those for the comparison between the eight radiologists and the final orthopaedic diagnosis. The scores for the collateral ligament were equally 'poor' in both comparisons.

In terms of satisfaction with image quality, although there were some differences of opinion

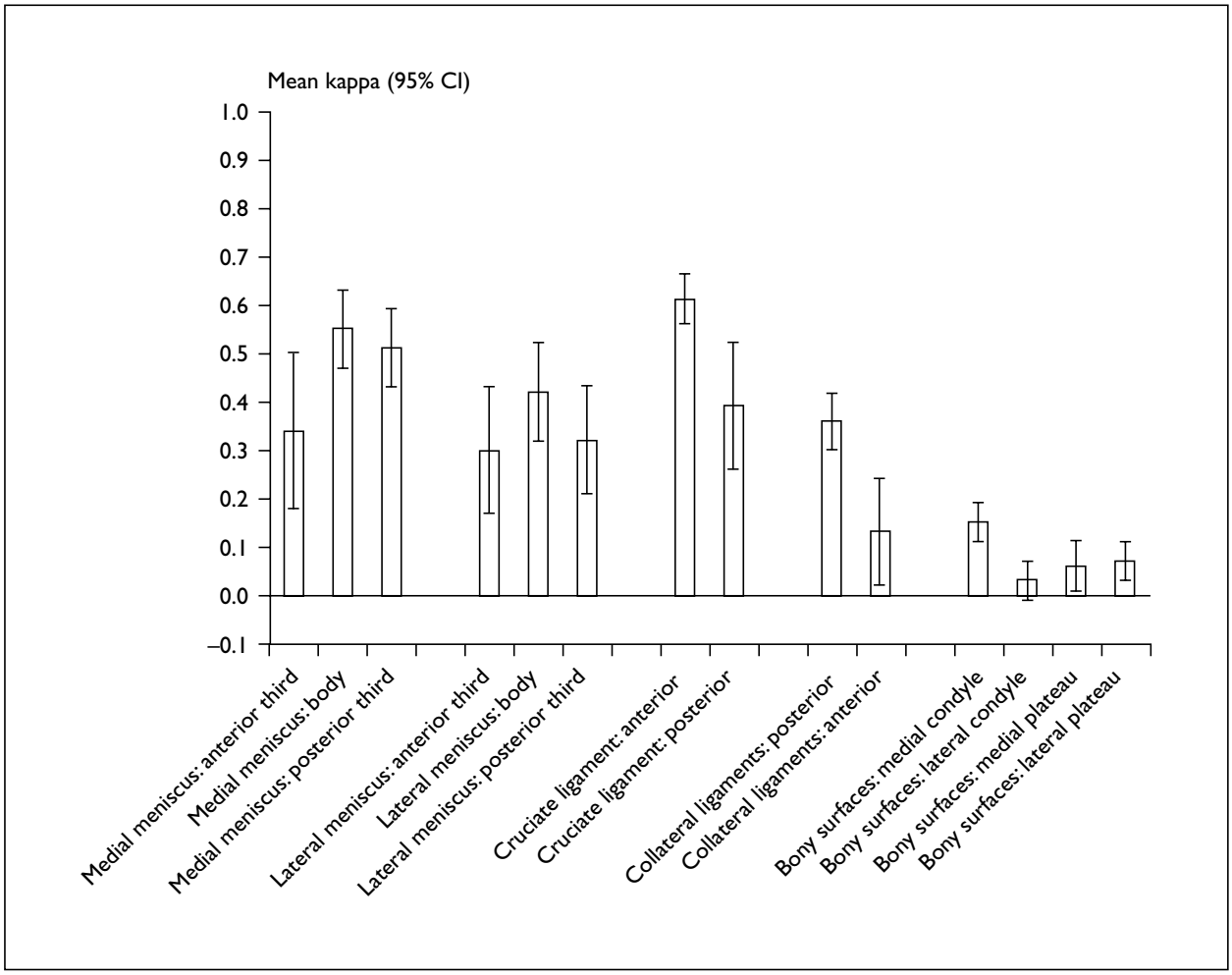


FIGURE 14 *Average levels of agreement between the final reference diagnosis and the diagnoses by the eight radiologists*

between viewers, the radiologists were satisfied with 90% of images overall and dissatisfied with 8% (with the assessment of satisfaction not given for 2% of examinations). The average time taken to view the MRI examinations and complete the proforma was 4 minutes, although this varied between radiologists (*Table 14*).

TABLE 14 *Time taken to view MRI images for each patient*

Viewer	Number of MRI examinations	Mean (SD) time (minutes)
1	80	5.48 (0.97)
2	78	3.03 (1.14)
3	79	1.99 (0.88)
4	78	2.21 (0.92)
5	80	10.13 (1.12)
6	80	3.99 (1.26)
7	80	3.87 (1.18)
8	80	2.67 (1.27)

Discussion

The original MRI images (which were produced on film) were viewed by all radiologists and, thus, the radiologists were all provided with exactly the same information. Therefore, none of the variation in interpretations can be explained by inter-image variations.

The MRI images included in this study had been produced according to the protocols of the specialist centre. The protocols in the DGHs may not have been identical and, thus, the images may have varied from those that the radiologists in the DGHs normally used. This may have accounted for some of the differences in interpretation of the images between radiologists. All radiologists were asked about their satisfaction with the quality of the images, and, although they were satisfied with the quality of 90% of the images, there was dissatisfaction with the quality of 8%. However, dissatisfaction was not only noted by DGH radiologists, but also by the radiologist in the specialist

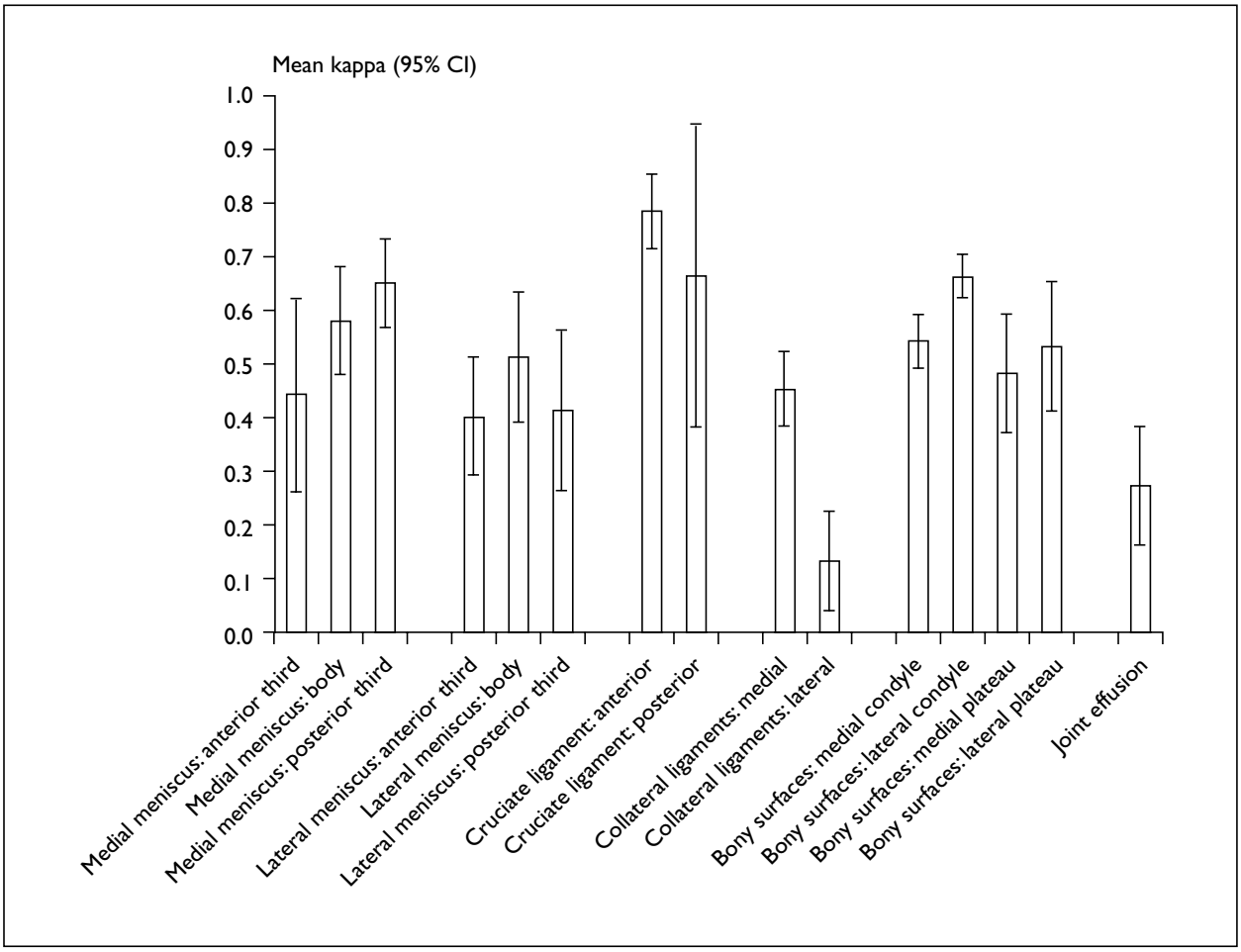


FIGURE 15 *Average levels of agreement between the radiologist in the specialist centre and the seven DGH radiologists*

centre, in whose department the examinations were undertaken. The specialist radiologist expressed dissatisfaction with 9% of the images.

The radiologist in the specialist centre made a report on each MRI examination soon after the examination was undertaken and this was subsequently coded using the study pro-forma. Therefore, it is possible that the specialist radiologist had additional clinical information about the MRI examination or about the patient, which were not available to the other radiologists. However, the accuracy of the radiologist in the specialist centre was no better than the average for all DGH radiologists.

The results of this study are in agreement with other studies, which have shown that MRI is sensitive for the diagnosis of medial meniscus tears and injuries to the ACL.[13,73,74] The results also agree with a study which showed that MRI has lower sensitivity for bony surfaces.[15]

Fischer and colleagues studied MRI investigations of the knee that were undertaken at four centres and reported by the local radiologists.[73] They found differences in the true-positive, true-negative, false-positive and false-negative results between the centres for the medial meniscus and the ACL, but not for the lateral meniscus and the posterior cruciate ligament. In the study reported in this chapter, it was found that the radiologists' scores and the final orthopaedic diagnosis were very similar for both the medial and the lateral menisci and the anterior and posterior cruciate ligaments. Fischer and colleagues found that one centre performed consistently less well compared with the other three centres, but could not determine whether this was due to the radiologist or the MRI unit and the images. Our study, which used the same images all produced by the same MRI unit, found that diagnostic accuracy was similar for all radiologists.

Conclusion

This study demonstrated that, in general terms, the diagnoses by the eight consultant radiologists

were very similar to the reference standard diagnoses. In addition, radiologists in DGHs provided similar interpretations of knee MRI images as a radiologist at a specialist centre.

The areas of the knee for which these findings did not hold are the lateral collateral ligament and the bony surfaces, where agreement tended to be poor.

Chapter 7

The value of the diagnostic and therapeutic impact of knee MRI: a stated preference survey

Introduction

This chapter reports two empirical conjoint measurement exercises that have been used as the vehicle to estimate the value that potential patients place on the possible diagnostic and therapeutic 'impacts' of knee MRI as identified in chapter 4. The first was a pilot study undertaken principally to provide information on some of the key factors in the design and conduct of a more definitive main study. However, the pilot study was also used to address some methodological issues that are of more general interest in the use of this form of analysis in health economics. (Full details of the pilot study are reported in Bryan and colleagues, 1998.[75]) The next section reports the pilot phase of the research, and is followed by a full description of the main conjoint exercise.

Pilot conjoint study

Methods

Attribute identification

The first stage of the pilot conjoint study was to define the attributes of interest. Four attributes were selected:

- the probability of receiving an arthroscopy
- the probability of the knee problem being completely resolved
- the time from initial hospital visit to the end of treatment
- cost.

One of the key attributes was the 'treatment path', defined in terms of the probability of receiving surgery. Another important attribute included in the exercise was health outcome, that is, the extent to which the knee problems were resolved. This attribute was defined in terms of the probability of the knee problem being 'completely resolved' and allowing the patient to undertake all normal activities without pain. Respondents were told that if their knee problem had not been completely resolved then the knee would be periodically painful, especially after sporting activities, such as rugby or football.

The process of being referred for an MRI scan might imply a delay in decision-making regarding the management and treatment of the problem.* If the scan indicated that an arthroscopy was required then the patient may be put on the waiting list later than they would have been had the surgeon opted for arthroscopy rather than MRI initially. In many parts of the country, access to MRI scanners is limited, especially for patients with knee injuries, and waiting times for MRI can be several weeks. It was, therefore, considered appropriate to include an attribute that described time from initial hospital attendance to the end of treatment.

Since there is limited access to MRI under the NHS in many parts of the country, many patients with knee injuries do not have access to MRI unless they are willing to purchase the scan privately. Given this reality, a fourth attribute of cost was considered. The exercise was designed such that the cost attribute was linked to the MRI scan and respondents were told that in some situations they would only be able to obtain an MRI scan if they paid for it privately. Respondents were asked to imagine a situation where knee MRI was not covered by any health insurance policy they may have and the cost would, thus, be an 'out-of-pocket' payment required when the MRI scan was undertaken. Including the cost attribute also allowed for the possibility of indirectly estimating patients' willingness-to-pay.

Level assignment

The preference elicitation approach adopted in this study imposed some restrictions on attribute

* This was not seen in the trial (see chapter 4) since the Ethics Committee required that trial patients were not disadvantaged in terms of prompt access to surgery as a result of participation in the study.

levels. An outline of the approach is, therefore, given at this point before the level assignment is described in detail. A discrete-choice paired-comparisons approach to preference elicitation was used on the basis that it represents a relatively simple and commonly encountered task for respondents and has a strong theoretical basis.[76] All choices were between two labelled scenarios: 'conventional treatment' (i.e. surgery) and 'MRI'.

An important consideration in level assignment was the inevitable limitation of each respondent's capacity to discriminate between alternatives, and the number of attribute levels was, therefore, restricted. The attribute levels chosen for the pilot study are detailed in *Table 15*.

For scenarios labelled as conventional treatment, there was only one level on the treatment attribute: patients always received surgery. The levels on the treatment attribute for the MRI alternative were chosen using data from published sources.[77,78] Respondents were told that for situations where MRI does not lead to therapeutic arthroscopy, management would be conservative, involving a course of physiotherapy.

The levels for the attribute time in the treatment process and health outcome were selected on the basis of advice from orthopaedic surgeon colleagues. For the arthroscopy alternative, there was only one level on the cost attribute: zero cost. The cost levels for the MRI alternative were set on the basis of published data by Birch and colleagues[78] and data supplied by the finance department of the hospital involved in the MRI knee trial (Kent and Canterbury Hospital).

Scenario presentation and preference elicitation

The next stage in this conjoint exercise was for selected hypothetical scenarios, with different combinations of attributes and levels, to be presented to respondents in order for their preferences to be determined. Given that there were four attributes, three of which had four levels and one of which had three, there were a total of 192 alternative scenarios with different combinations of attribute levels that could have been presented. It was, therefore, impossible to use a 'full factorial design' in which all scenarios were presented separately. A fractional factorial design was, therefore, used that allowed only main effects to be investigated, and 16 scenarios to present to respondents were selected using the software package MINT (Hague Consulting Group, 1990).† A total of 16 pairs of scenarios were selected for presentation, again using the MINT software package.

Conjoint questionnaire

The study used a self-completion questionnaire to elicit preferences. Eight choices were included in each questionnaire in order to avoid respon-

TABLE 15 *Attributes and levels used in the pilot study*

Attributes	Levels
Treatment	100% chance of requiring an arthroscopy 90% chance of requiring an arthroscopy 80% chance of requiring an arthroscopy 70% chance of requiring an arthroscopy
Time from initial consultation to end of treatment process	6 weeks 12 weeks 18 weeks 24 weeks
Resolution of knee problem	90% chance that knee problem is completely resolved 70% chance that knee problem is completely resolved 50% chance that knee problem is completely resolved
Total cost of MRI to the patient (i.e. cost only appears in the MRI scenarios)	zero £50 £100 £200

† The process of selecting the 16 scenarios for presentation and their pairing was undertaken by the professional market researchers Accent Marketing and Research.

dent fatigue, and, thus, two versions of the questionnaire (A and B) were developed in order to allow all 16 choices to be presented. Two dominated choices were included in each questionnaire. The information provided to respondents and an example of the choices is shown in appendix 7.

The questionnaire included an information sheet that outlined the key features of the technologies being compared. As a warm-up exercise, the first section asked respondents to rate the importance of the attributes and identify any other factors that would be important to them in the situation of having a serious knee injury. The second section included the eight conjoint choices. Respondents were asked to tick one box for each choice indicating a preference either for the conventional treatment or for MRI. In the third section, respondents were asked to provide contextual information concerning:

- age and sex
- whether they had ever had a knee injury
- whether they had ever had either an arthroscopy or an MRI scan
- whether they believed all healthcare should be provided without charge
- the time they devoted to sports activities per week
- the level of their sporting involvement (regional or international levels)
- whether they felt that they had sufficient information to sensibly answer the conjoint questions.

Sample and data collection

The target population chosen for this study was undergraduate students at Brunel University College taking a degree course in Sports Science. It was expected that most of the students on this course would be actively involved in sporting activities, a group amongst whom knee injuries are common. The exercise was 'framed' as an investigation into what services and facilities should be provided in specialist sports injury clinics. The students were asked to remain behind at the end of a lecture to complete a questionnaire.

Data analysis

Each respondent considered eight choices and thus provided up to eight data points. Therefore, the conjoint data are not independent. In response to the non-independence of data points, a random-effects probit model was estimated where the conjoint responses were modelled using an additive functional form as follows:

$$y_{ijq} = \alpha_0 + \alpha_1(a_j - a_i) + \alpha_2(t_j - t_i) + \alpha_3(k_j - k_i) + \alpha_4(c_j - c_i) + v_{ijq} + u_q$$

where i is the left-hand scenario within a choice; j is the right-hand scenario within a choice; $ij = 1, \ldots, 8$ (the number of conjoint choices posed); $q = 1, \ldots, n$ (the number of respondents to the survey); $y = (RU_j - RU_i)$, i.e. the difference in random utility (RU) between the two scenarios presented within a choice; a, t, k, c are the levels for the treatment, time, knee problem resolution and cost attributes, respectively; $\alpha_0 \ldots \alpha_4$ are the model coefficients; v_{ijq} is the random error term due to differences amongst observations; and u_q is the disturbance due to differences amongst respondents resulting from measurement error.

Results

Sample and descriptive data

All 52 students who were asked to complete the questionnaire responded. Not surprisingly, the majority of students were male (56%) and in their early twenties (mean age = 20 years). Almost half of those in the sample had suffered a knee injury that required the attention of a doctor. About 15% had received an MRI scan and a similar percentage had received arthroscopic surgery on their knee. All respondents undertook some sports or exercise activities each week and over 80% devoted more than 5 hours/week to such activities.

Given that each respondent was asked to consider eight choices, the maximum number of conjoint data points that would have been available had all respondents fully completed every questionnaire was 416. The number of choices where the respondent expressed a clear preference for one or other option was 387 (93.0%).

The majority of all respondents indicated that the attributes chosen for inclusion in the study were either 'important' or 'very important' (*Table 16*). A total of 18 respondents (35%) indicated that there were other factors that would be important to them in the situation of suffering a knee injury (see *Box 1*).

Conjoint data

Table 17 reports the results of the random-effects model. The $\prod^2$ statistic compared a 'constant-only' model with the specified model, and the reported significance level indicated whether, taken together, the coefficients were significant. The model fit can also be determined from the predictive power, measured by the percentage of choices correctly predicted.

TABLE 16 *Importance of attributes in pilot study (number of respondents)*

Attribute	Very important	Quite important	Of little importance	Of no importance
Avoidance of delays	46 (88.5%)	5 (9.6%)	1 (1.9%)	0 (0.0%)
Avoidance of surgery	9 (17.3%)	25 (48.1%)	17 (32.7%)	1 (1.9%)
Avoidance of payment	18 (34.6%)	21 (40.4%)	9 (17.3%)	4 (7.7%)
Resolution of knee problem	49 (94.2%)	3 (5.8%)	0 (0.0%)	0 (0.0%)

BOX 1 Other important factors not included in the pilot study

- Short recovery time following treatment
- Avoiding pain or scarring
- Information about the problem/what treatment involves
- Advice on recuperation/good postoperative care
- High degree of confidence in doctors/surgeons
- Avoiding recurrence of the problem
- Correct diagnosis
- Good attitude of hospital staff
- Flexibility of time of treatment, e.g. out-of-season
- Avoiding problems in later life
- Avoiding time off work
- Location of operation/treatment
- Comfortable waiting area
- Care provided by the NHS

The attributes relating to arthroscopy, time involved in treatment process and solving of the knee problem were all highly significant. This indicates that, in general, all three factors were important to respondents when making their choices. The signs on these three attributes were as expected. The negative sign on arthroscopy indicated that, other things being equal, as the probability of requiring an arthroscopy increased under the MRI option, respondents were more likely to choose the conventional treatment option. The negative coefficient on the time involved in the treatment process indicated that, other things being equal, if treatment time under the MRI option was shorter, respondents were more likely to choose MRI. The positive coefficient on the resolution of the knee injury indicates that, other things being equal, if the probability of the knee problem being resolved under the MRI option was lower, respondents were more likely to choose

TABLE 17 *Pilot study probit models*

	Random-effects probit model		Random-effects probit model (excluding non-traders)	
Attributes	**Coefficient**	**95% CI**	**Coefficient**	**95% CI**
Arthroscopy	–2.952[a]	–4.557 to –1.347	–2.341	–5.128 to 0.445
Time involved in treatment process	–0.056[a]	–0.078 to –0.033	–0.057[a]	–0.094 to –0.021
Resolution of the knee injury	6.495[a]	5.368 to 7.622	4.733[a]	2.930 to 6.536
Out-of-pocket payment	–0.001	–0.005 to 0.002	–0.005	–0.011 to 0.001
Constant	–0.724[a]	–1.243 to –0.204	–0.687	–1.496 to 0.122
ρ	0.077		0.028	
n (data)	413		120	
n (groups)	52		15	
Mean group size	7.94		8	
$\prod^2$	144.56		26.85	
p	0.000		0.000	
Predictions:				
y = 1	74%		58%	
y = 0	75%		78%	

[a] $p < 0.01$

conventional treatment. The cost attribute was not significant indicating that respondents did not consider the treatment cost, within the range specified in the study, to be an important factor when making their choices. The conventional interpretation of the constant term in such models is the propensity to choose option *j* over option *i* (or the propensity to choose left rather than right) with all other things remaining equal. Given that the conjoint choices in this exercise involved labelled scenarios (i.e. conventional treatment and MRI), the constant term had a more significant role: it indicated the underlying preference for one of the technologies. The negative sign suggested a general preference for the conventional treatment option (i.e. arthroscopy).

The ratio of the attribute coefficients in the probit models provided an indication of the average marginal rate of substitution (MRS) between attributes. These are presented in *Table 18*. For example, the results indicated that, on average, respondents were willing to exchange a fall of 10% in the chance of the knee problem being resolved for an increase of 22% in the chance of an arthroscopy being avoided (i.e. resolution of the knee injury:arthroscopy = 2.2). Similarly, on average, respondents were willing to exchange a fall of 10% in the chance of the knee problem being resolved for a reduction of 12 weeks in the duration of treatment. The use of the model constant in a similar way revealed the change in attribute levels that respondents were willing to see, on average, for MRI to be used. For example, respondents were, on average, prepared to see a lower chance of the knee problem being resolved (by 11.2%) in order to ensure the use of MRI.

Test of lexicographic preferences

A total of 37 respondents (71%) displayed lexicographic preferences, most choosing the option that represented the best chance of resolving the knee problem. Given the high proportion of lexicographic respondents, the random-effects probit model was re-estimated excluding data from all lexicographic respondents. In broad terms, the key results were similar. The CIs around the coefficient estimates were wider, as one would expect, given that fewer data points were used in the re-estimation of the models. Since the majority of lexicographic responses related to resolving the knee problem, it was not surprising that the coefficient estimates for that attribute were affected most by dropping data from lexicographic respondents. However, the re-estimated coefficients for the resolution of the knee injury remained significantly different from zero and positive, and the coefficient estimates for all other attributes and for the constant term were within the 95% CIs estimated for the full data model. However, one notable difference was that arthroscopy was no longer significant, even though the value of the coefficient was very similar to that estimated in the full data model. This finding may simply reflect the smaller sample and associated wider CIs, or may indicate that the trading subgroup were not concerned by the diagnostic and therapeutic impact of MRI (i.e. the chance of avoiding surgery).

TABLE 18 *Pilot study MRSs*

MRS between:	MRS
Chance of avoiding arthroscopy and cost	None[a]
Time in treatment process and cost	None[a]
Chance of knee problem resolution and cost	None[a]
Use of arthroscopy and cost	None[a]
Time in treatment process and chance of avoiding arthroscopy	1.9%: 1 week
Chance of knee problem resolution and chance of avoiding arthroscopy	2.2% (arthroscopy avoidance): 1% (knee problem resolution)
Use of arthroscopy and chance of avoiding arthroscopy	24.5%: use of arthroscopy
Chance of knee problem resolution and time in treatment process	1.2 weeks: 1%
Use of arthroscopy and time in treatment process	12.9 weeks: use of arthroscopy
Use of arthroscopy and chance of knee problem resolution	11.2%: use of arthroscopy

[a] *On average, respondents were unwilling to exchange variation in cost for variation in the other attributes, using the attribute levels applied in the pilot study*

Main conjoint study

Introduction

One of the main findings of the pilot study was that many respondents did value the diagnostic and therapeutic impacts of knee MRI. Given the importance of this conclusion for the evaluation of diagnostic technologies, one of the key objectives of the main conjoint analysis study was to establish the robustness of this finding using a larger sample.

The pilot study also highlighted important design issues. The relatively high proportion of lexicographic responders found in the pilot study was undesirable, since the level of information provided by such respondents in terms of MRSs between attributes is inevitably limited. The lexicography problem in the pilot study might have been avoided if the intervals between the attribute levels had been larger. In theory, one can almost always ensure trading by setting extreme intervals between levels. However, in doing so, the realism of the stated preference exercise is potentially diminished. If respondents will only trade at levels that are 'not implementable' then the results of the conjoint exercise are of limited value in public policy terms.

The cost attribute used in the pilot study used levels for cost chosen to reflect the prices patients might actually face if they were to purchase an MRI scan privately. This was thought to retain realism in the exercise but had the disadvantage that any willingness-to-pay estimates would be constrained to that range. Such design trade-offs are an inevitable feature of conjoint analysis studies. It is interesting to note that other conjoint studies of healthcare interventions have found cost to be a significant attribute.[79,80] The fact that the cost attribute was not significant in the pilot study indicated only that, within the range of costs chosen, the choices were principally influenced by the other attributes. The use of a wider range for cost, either with larger intervals between levels or more levels, might have given a different result, but equally might have been viewed as unrealistic by respondents. This issue was also addressed in the main study.

Methods

Study design

The framework for the main conjoint analysis study was retained as a discrete choice comparison of MRI and arthroscopy. The labelling of the alternatives as MRI and arthroscopy was also preserved. The levels of some of the attributes were amended in light of the problems encountered in the pilot study. As indicated above, the high proportion of lexicographic respondents and the fact that the cost attribute was not significant were thought possibly to reflect the selection of inappropriate levels. This issue was explored by design changes in the attribute levels for three attributes: the probability of requiring arthroscopy, the probability of resolution of the knee problem and the cost to the patient. The knee problem resolution attribute represented by far the largest problem and thus the number of levels was increased from three to four and the interval between levels was reduced (from 20 to 10%). For the cost attribute, the levels between attributes were increased, although the number of levels remained as four in order to limit the escalation in the number of scenarios included in the questionnaire. The revised levels were selected to represent a more challenging range to respondents, forcing them to consider whether MRI was a realistic option for them, whilst attempting to retain enough reality in the scenarios such that respondents were sufficiently engaged in the exercise. This is a common difficulty in conjoint study design and requires judgement by the researcher. The time attribute levels were not changed since this attribute represented the smallest problem in terms of lexicographic respondents (i.e. only 4 out of 134). For the arthroscopy attribute, the number of levels was reduced from four to three and the interval between the levels was increased. This was done in order to establish whether the finding in the pilot study (i.e. the significance of the arthroscopy attribute in determining choices) was repeated in the main study. The revised attribute levels are shown in *Table 19*.

Again, there were 192 alternative scenarios available for presentation, and an orthogonal fractional factorial main-effects design, without interaction terms, was used. A total of 20 scenarios were chosen using MINT: ten relating to arthroscopy and ten relating to MRI. The software also provided the pairings of scenarios for the choices, and a total of 12 choices were selected for presentation.[‡]

As in the pilot study, self-completion questionnaires were used to collect data, although, in the

[‡] The process of scenario selection for presentation and the pairing of scenarios was again undertaken by the professional market researchers Accent Marketing and Research.

TABLE 19 *Attributes and levels used in main study*

Attributes	Levels
Treatment	100% chance of requiring an arthroscopy 80% chance of requiring an arthroscopy 60% chance of requiring an arthroscopy
Time from initial consultation to end of treatment process	6 weeks 12 weeks 18 weeks 24 weeks
Resolution of knee problem	90% chance that knee problem is completely resolved 80% chance that knee problem is completely resolved 70% chance that knee problem is completely resolved 60% chance that knee problem is completely resolved
Total cost of MRI to the patient	zero £200 £400 £600

main study, all 12 choices were included in each questionnaire. This increase in choices was made since no obvious 'fatigue effects' were found in the pilot study.

The questionnaire was very similar in appearance to the pilot study questionnaire. Respondents were again also asked to provide information on their age and sex, whether they had ever had a serious knee injury, their experience of the technologies in question and their level of participation in sport.

The target population chosen for the main study was undergraduate students studying for degrees in Sports Science or trainee sports teachers. Data were collected from five universities in the UK: Brunel, Exeter, Kingston, Staffordshire and Birmingham. The data collection process was the same as in the pilot study: questionnaires were completed either at the end or at the beginning of a lecture following a short introductory talk.

Hypotheses and data analysis

The central hypotheses tested in the main conjoint study were that patients with knee injuries would:

- prefer to avoid surgery
- prefer to avoid longer treatment processes
- prefer a higher chance of the knee problem being resolved
- prefer a lower out-of-pocket payment.

Random-effects probit models were used to analyse the conjoint data, given that each respondent provided multiple (up to 12) observations. The same model specifications that were used in the pilot study were employed in the main study.

The baseline random-effects probit model was estimated using all data, regardless of whether respondents exhibited lexicographic preferences. This model was defined as 'baseline A'. Given the current debate concerning the appropriateness of including data from respondents who are unwilling to trade in the conjoint modelling (e.g. see Ryan and Hughes[81]), a second baseline model was estimated using only data from trading respondents, which was defined as 'baseline B'.

Results

Sample and descriptive data

Data were collected from 585 students in total: 188 from Staffordshire, 163 from Exeter, 149 from Brunel, 69 from Birmingham and 16 from Kingston. Given the nature of the data collection exercise, essentially with a 'captive' audience, the response rate was, in essence, close to 100% of those students who attended the lectures. The majority of respondents were male (58%) and within an age range expected for undergraduate university students (mean age = 20.4 years, median = 19, range 18–40). *Table 20* provides further descriptive information about the respondents.

Data on the responses to the initial questions about the importance of each attribute are shown in *Table 21*. Whilst the majority of all respondents indicated that the four attributes chosen for inclusion in the study were either 'important' or 'very important', the proportion of respondents

TABLE 20 *Main study descriptive information (number of respondents replying 'Yes')*

Question	Number (%) of respondents
Have you ever had a serious knee injury?	235 (40%)
Have you ever had an MRI scan?	40 (7%)
Have you ever had an arthroscopy?	51 (9%)
Have you represented the region/county in the last 12 months?	316 (54%)
Have you represented the country in the last 12 months?	62 (11%)
Do you undertake more than 5 hours exercise/week?	441 (75%)
Do you think all healthcare should be provided free of charge?	447 (76%)
Was there enough information given in the questionnaire?	517 (88%)
Both knee injury + MRI	30 (5%)
Both knee injury + arthroscopy	44 (8%)
Both knee injury + regional/international athlete	146 (25%)
Both knee injury + more than 5 hours of exercise/week	180 (31%)
Both MRI + arthroscopy	8 (1%)
Both MRI + regional/international athlete	28 (5%)
Both MRI + more than 5 hours of exercise/week	37 (6%)
Both arthroscopy + regional/international athlete	33 (6%)
Both arthroscopy + more than 5 hours of exercise/week	43 (7%)
Some of the totals do not sum to 585 because of missing data	

TABLE 21 *Importance of attributes to main study respondents (number of respondents)*

Attribute	Very important	Quite important	Of little importance	Of no importance
Avoidance of delays	487 (83.4%)	94 (16.1%)	3 (0.5%)	0 (0.0%)
Avoidance of surgery	128 (22.0%)	291 (50.1%)	145 (25.0%)	17 (2.9%)
Avoidance of payment	282 (48.5%)	228 (39.2%)	62 (10.7%)	9 (1.5%)
Resolution of knee problem	563 (96.4%)	19 (3.3%)	2 (0.3%)	0 (0.0%)

indicating that the attribute was 'very important' was lowest for the avoidance of surgery attribute. A total of 133 (23%) respondents indicated that, in the situation of having a knee injury, there were other factors that would be important to them (see *Table 22*).

A total of 104 (18%) respondents revealed lexicographic preferences and were unwilling to trade (*Table 23*). This proportion was much smaller than that found in the pilot study and indicated that the adjustments made to the attribute levels were effective. Most lexicographic respondents ($n = 54$) were unwilling to trade cost and always chose the lowest cost alternative. This result was in contrast to the pilot study where the vast majority of lexicographic respondents were unwilling to trade on the knee problem resolution attribute. In addition to the 104 lexicographic respondents, a further 19 respondents consistently chose one of the two technologies, regardless of the attribute levels: 16 always chose the surgery option and three always chose the MRI option.

Conjoint models

Table 24 reports the results of the baseline random-effects probit model using all data (baseline A). Both the $\prod^2$ statistic and the predictive power of the model suggested a reasonable fit for the baseline model. The attributes relating to cost, time involved in the treatment process and solving of the knee problem were all highly significant, indicating that, in general, variation in these three factors influenced respondents when making their choices. The signs on these three attributes were as expected, and were the same as

TABLE 22 *Other important factors not included as main study attributes*

Other factors	Number of respondents
Short recovery time following treatment	53
Good follow-up care, including physiotherapy	21
Information about the problem/what treatment involves	13
Avoiding pain or scarring	13
Care by a specialist clinical team	8
Care provided locally/avoid travel problems	7
Avoid side-effects	6
Correct diagnosis	5
Good attitude of hospital staff	3
Patient given treatment choice	1
Avoid recurrence of problem	1
Able to return to work	1
Avoid drugs if possible	1

TABLE 23 *Number of main study respondents revealing lexicographic responses*

Attribute	Number (%) of lexicographic respondents
Avoidance of delays	13 (2.2%)
Avoidance of surgery	3 (0.5%)
Avoidance of payment	54 (9.2%)
Resolution of knee problem	34 (5.8%)

TABLE 24 *Main study random-effects probit model (baseline A)*

Attributes	Coefficient	95% CI
Arthroscopy	0.148	–0.119 to 0.415
Time involved in treatment process	–0.029[a]	–0.033 to –0.024
Resolution of the knee injury	3.899[a]	3.662 to 4.136
Out-of-pocket payment	–0.003[a]	–0.003 to –0.003
Constant	0.325[a]	0.230 to 0.419
ρ	0.139	
n (data)	6970	
n (groups)	585	
Mean group size	11.91	
$\prod^2$	2081.64	
p	0.00	
% correct predictions:		
y = 1	55%	
y = 0	79%	

[a] $p < 0.01$

TABLE 25 *Main study MRSs*

MRS between:	Baseline A	Baseline B
Chance of avoiding arthroscopy and cost	None[a]	None[a]
Time in treatment process and cost	10.78: 1 week	13.61: 1 week
Chance of knee problem resolution and cost	14.55: 1%	13.77: 1%
Use of arthroscopy and cost	121.17: use of MRI	164.25: use of MRI
Time in treatment process and chance of avoiding arthroscopy	None[a]	None[a]
Chance of knee problem resolution and chance of avoiding arthroscopy	None[a]	None[a]
Use of arthroscopy and chance of avoiding arthroscopy	None[a]	None[a]
Chance of knee problem resolution and time in treatment process	1.34 weeks: 1%	1.01 weeks: 1%
Use of arthroscopy and time in treatment process	11.23 weeks: use of MRI	12.06 weeks: use of MRI
Use of arthroscopy and chance of knee problem resolution	8.33%: use of MRI	11.92%: use of MRI

[a] *On average, respondents were unwilling to exchange variation in the chance of avoiding arthroscopy for variation in the other attributes, using the attribute levels applied in the main study*

those found in the pilot study. For example, the interpretation of the negative sign on out-of-pocket payment was that, other things being equal, as the cost associated with MRI increased, the respondents were less likely to choose MRI. However, the attribute representing diagnostic and therapeutic impact, i.e. the avoidance of knee surgery (arthroscopy), was not significant indicating that respondents did not consider the probability of avoiding arthroscopy, within the range specified in the study, to be an important factor.

The MRSs between attributes from the main study models are shown in *Table 25*. For example, on average, respondents were willing to see an increase in cost of just over £10 in return for a 1-week reduction in the total treatment time. Similarly, respondents were willing, on average, to accept an increase in cost of approximately £145 in return for an increase of 10% in the probability of the knee problem being resolved.

Another important difference from the pilot study was the sign on the constant coefficient. For the main study, the coefficient was positive indicating a systematic tendency to choose the right-side option, which was always MRI (in both the pilot and main studies). The common interpretation of this finding, as discussed above, is the expression of an underlying preference for MRI, all other things being equal. In contrast, the pilot study model had a negative constant coefficient.

The baseline model using only data from respondents who were willing to trade (baseline B) is shown in *Table 26*. The $\prod^2$ statistic and the predictive power of the model suggested a slightly better fit for the baseline B model compared to baseline A. This was not surprising given that lexicographic respondents were excluded, whose responses, by definition, are not well represented by a probit model that assumes some trading between attributes. The key results were similar to the baseline A model, which was based on all data. In line with prior expectations, the attribute coefficients were generally larger when lexicographic respondents were excluded, indicating the obvious fact that the responses of trading respondents are more sensitive to variation in attribute levels. The CIs around the coefficient estimates in baseline B were slightly wider, as one would expect, given that fewer data points were used in their estimation. The attributes that were significant in the baseline A model remained significant in the baseline B model and the sign on each coefficient was the same. The one attribute where the difference reached statistical significance was treatment time.

Discussion

An important finding of the pilot study was that respondents were, on average, willing to trade between the diagnostic and therapeutic impact of knee MRI (i.e. avoidance of surgery) and health outcome (i.e. resolution of the knee problem), indicating that such impacts were considered valuable consequences from the use of MRI. However, the pilot study results should be viewed

TABLE 26 *Main study random-effects probit model (baseline B)*

Attributes	Coefficient	95% CI
Arthroscopy	0.294	–0.039 to 0.630
Time involved in treatment process	–0.042[a]	–0.048 to –0.037
Resolution of the knee injury	4.272[a]	3.994 to 4.549
Out-of-pocket payment	–0.003[a]	–0.003 to –0.003
Constant	0.509[a]	0.396 to 0.622
ρ	0.079	
n (data)	5724	
n (groups)	481	
Mean group size	11.90	
$\prod^2$	1659.71	
p	0.00	
% correct predictions:		
y = 1	59%	
y = 0	81%	

[a] p < *0.01*

with caution for two specific reasons: firstly, the sample was small and thus prevented the use of data segmentation and, secondly, the proportion of respondents who were not willing to trade between attributes was large. In response, the main study sought to recruit a much larger sample, which was successfully achieved, and adjustments were made in the design of the conjoint exercise, as described above, to avoid a similarly high level of lexicographic response. The main study was also successful in the second objective: the percentage of lexicographic respondents fell from 71% in the pilot study to only 18% in the main study. Whilst it could be argued that this was only achieved at the cost of using levels for some attributes that have little policy relevance (for example, it is most unlikely that a patient would face a price of £600 for a knee MRI scan), there seems little reason to believe that this had an adverse effect on respondent engagement (for example, the time taken to complete the questionnaires was similar between the main and pilot studies).

Given the larger sample and the re-design of the questionnaire in light of pilot experience, greater confidence can be placed in the main study results. On average, the diagnostic and therapeutic impact of knee MRI was **not** an important concern to respondents in the main study: the coefficient on the arthroscopy attribute was not significantly different from zero, indicating that, within the range specified, choices were not influenced by variation in the probability of avoiding surgery. There were two other important differences between the pilot and main studies. Firstly, the cost attribute changed from not being significant to being significant, indicating that the study re-design in relation to this attribute was successful and that respondent choices **were** sensitive to the level of cost, as others have demonstrated.[79,80] Secondly, the sign on the constant coefficient changed from negative (in the pilot study) to positive (in the main study). The interpretation of this is a switch from an underlying preference for arthroscopy to a preference for MRI, all other things being equal. This represents a notable difference, especially given the strength of the preferences shown by the MRSs reported in *Tables 18* and *25*. Bryan and Parry[82] provide a discussion of issues concerning the stability and interpretation of the constant coefficient. For the other two attributes, treatment time and knee problem resolution, the nature of the preferences identified was similar in both studies.

The conjoint exercises reported here both re-analysed the data excluding respondents with lexicographic responses. The full data model provided an estimate of the **aggregate** MRSs between attributes for the whole group, and is most relevant to economic evaluation because the decision-making context is the allocation

of scarce healthcare resources for pre-defined groups (e.g. the general population in the UK or subscribers to a health insurance plan in the USA). The model derived using only data from trading respondents (as advocated by some researchers, e.g. Ryan and Hughes[81]) provided estimates of MRSs for the subgroup of respondents who appeared willing to make trade-offs between attributes at the levels indicated.[81] The views of this subgroup are of interest in two respects. Firstly, if they represent a discrete subgroup with other common characteristics (e.g. age, ethnic group, etc.) then it becomes possible for their different views to result in policy variation, although a better approach to identifying such groups would be to segment using known characteristics and not by whether they appeared to trade between attributes. Secondly, repeating the analyses using only data from the subgroup of trading respondents provides one possible approach to establishing the robustness of the average results obtained for all respondents. This was the approach used in this chapter.

An important consideration when interpreting the results from this conjoint study is that the data were drawn from a sample of university students and, therefore, concerns exist around generalisability of the findings. This comment raises the central question of whose values are we interested in when considering resource allocation issues in a public healthcare system? There are legitimate claims both for citizen/public values (i.e. given that it is a public resource that is being allocated) and for patient or potential patient values (i.e. given that they have the experience of relevant factors). If the view is taken that we are only interested in the values of the general public, then the data reported here might be viewed as inconsequential. However, if legitimacy is given to the views of the users (and potential users) of the service, then the data from this chapter are relevant. The limitation is that the sample from which conjoint data were collected represents a subgroup of all patients presenting with knee problems in orthopaedic clinics. Therefore, whilst it is appropriate to be cautious in drawing policy conclusions on the basis of these data alone, they do nevertheless illustrate something interesting about preferences relating to a large subgroup of all patients presenting with knee injuries. The study reported here is one of the largest health-related conjoint studies to have been conducted.

Conclusion

The conjoint measurement studies reported here have demonstrated how preferences for a broad range of attributes associated with healthcare interventions, and especially those typically associated with diagnostic imaging technologies, can be measured using conjoint techniques. The central finding of importance for this report is that the diagnostic/therapeutic impact associated with the use of MRI for knee injuries was not highly valued. Respondents were unwilling to trade reductions in the chance of receiving surgery for variation in the levels on the other attributes. However, caution is required in drawing policy conclusions on the basis of these results alone, given that the data were collected from a sample of university students.

Chapter 8

Discussion and conclusions

Introduction

The research presented in this report considered the role of MRI in the diagnosis of knee abnormalities and injuries in a DGH setting. The principal objective was to identify whether the use of MRI in patients presenting with knee problems had a major impact on the clinical management of patients, whether it brought about an overall reduction in costs and whether it improved patient outcome.

In addition, the research:

- explored how diagnostic accuracy of the initial clinical investigation varied across clinicians (i.e. orthopaedic trainees, a consultant knee specialist and a consultant radiologist)
- considered the variability and diagnostic accuracy of interpretations of knee MRI investigations between radiologists from typical DGH settings
- measured the strength of preference for the potential diagnostic/therapeutic impact of knee MRI (i.e. the avoidance of surgery).

The main conclusions of the research in relation to each of these issues are briefly reiterated here, together with the key issues that are still to be resolved.

Principal findings

Trial results

In terms of patient health outcomes, the central finding was that there were no statistically significant differences between groups in all measures of health outcome, although a trend in favour of the no-MRI group was observed. However, the use of MRI was found to be associated with a positive diagnostic/therapeutic impact: a significantly smaller proportion of patients in the MRI group underwent surgery ($p = 0.001$). Overall, similar mean NHS costs for both the MRI and no-MRI groups were found, indicating that the increased cost associated with the use of MRI in all patients was offset in full by the reduced requirement for surgery.

Substudies

Investigation of diagnostic accuracy

The investigation of diagnostic accuracy of the initial clinical investigation of the knee provided data, which suggested that, when compared to orthopaedic trainees (44% correct diagnoses) or to radiologists reporting on an MRI scan (68% correct diagnoses), the accuracy rate was higher for knee specialists (72% correct diagnoses). Therefore, reliance on the results of an MRI scan without the necessary skilled clinical assessment may lead to the mismanagement of patients.

Investigation of the generalisability of results

This substudy demonstrated that, in general terms, radiologists in DGHs provided broadly comparable and consistently accurate interpretations of knee MRI images, and reports that were similar to a radiologist at a specialist centre. The areas of the knee for which these results did not hold were the lateral collateral ligament and bony surfaces, where agreement tended to be poor.

Investigation of preferences

The central finding for the conjoint measurement preference study was that, on average, the diagnostic and therapeutic impact of knee MRI was **not** an important concern to respondents. The coefficient on the arthroscopy attribute in the probit model was not significantly different from zero, indicating that, within the range specified, choices were not influenced by variation in the chance of avoiding surgery. However, caution should be exercised in generalising from this result, given that data were collected from a student sample.

Recommendations

Implications for healthcare

The evidence presented in this report lends support to the conclusion that the use of MRI in patients presenting at DGHs with chronic knee problems, in whom arthroscopy is being considered, does not increase NHS costs, is not associated with significantly worse outcomes and avoids surgery in a significant proportion of patients.

Recommendations for further research (in priority order)

1. The trial data reported here clearly demonstrated that the use of MRI in patients who had chronic knee problems reduced the need for surgery. However, the link between diagnostic processes and changes in health outcome is indirect and the finding of a no-MRI-related effect on health outcome may, therefore, be a consequence of the limited power available in this trial. Further research to confirm (or contradict) the findings of the trial data would be valuable.

2. The investigation of the diagnostic accuracy of the initial clinical investigations reported here involved comparison with a reference diagnosis obtained from the review of images and all patient notes by a panel of two members of the research team. The term 'gold standard diagnosis' was not used because the review panel was internal. It would be interesting to explore the extent to which our results would have been different had an external panel (i.e. clinicians outside the research team) been used. Given the importance of the quality of the MRI and clinical diagnosis to patient outcomes and resource utilisation, this represents one of the research priorities in this area.

3. The results from the preference study, indicating that the potential diagnostic/therapeutic impact of knee MRI was **not** highly valued, was a surprising and potentially very important finding. However, it may have been specific to the sample of respondents (i.e. sports science university students) and/or to the framing of the questions. It would be interesting to explore the extent to which these results vary when the general population or other patient populations are surveyed, and when questions are framed to include greater variation in attribute levels.

4. The focus for the trial-based aspects of this research was on a DGH and on patients presenting, typically, with chronic knee problems in whom arthroscopy was being considered. It is not possible to generalise from these results to other patient groups (e.g. those presenting with more acute knee injuries) or to other settings (e.g. specialist centres). Further trials would be required in order to answer such questions.

Acknowledgements

The authors would like to thank all those who advised and supported the research team at various stages throughout the project, in particular Roger Beech, Martin Buxton, Ian Russell, Cindy Billingham, Stephen Brealey and Robyn Jones. They are indebted to everyone who worked so hard on the ground (Kate Hampton, Scarlet McNally, Nigel Blackburn, John Bingham and Karen Arnold) and to Sue Cashmore who prepared the final manuscript. The authors are also very grateful to the three anonymous referees and the monograph series editors for their very helpful comments. In addition, the authors would like to thank all the patients and hospital staff who generously gave their time to help with this research. Finally, the views expressed are those of the authors, who are also responsible for any errors.

References

1. Thornbury JR. Clinical efficacy of diagnostic imaging: love it or leave it. *Am J Radiol* 1994;**162**:1–8.
2. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making* 1991;**11**:88–94.
3. Blackmore CC, Smith WJ. Economic analyses of radiological procedures: a methodological evaluation of the medical literature. *Eur J Radiol* 1998;**27**:123–30.
4. Fineberg HV, Bauman R, Sosman M. Computerized cranial tomography. *J Am Med Assoc* 1977;**238**:224–7.
5. Abrams HL, McNeil BJ. Computed tomography: cost and efficacy implications. *Am J Roentgenol* 1978;**131**:81–7.
6. Fineberg HV. Evaluation of computed tomography: achievement and challenge [editorial]. *Am J Roentgenol* 1978;**131**:1–4.
7. Thornbury, Kido DK, Mushlin AI, Phelph CE, Mooney C, Fryback DG. Increasing the scientific quality of clinical efficacy studies of magnetic resonance imaging. *Investigative Radiol* 1991;**26**:29–35.
8. Evens RG. Economic implications of a new technology installation: a CT model. *Am J Roentgenol* 1981;**36**:673–7.
9. Donaldson C. Using economics to assess the place of screening. *J Med Screen* 1994;**1**:124–9.
10. Miedzybrodzka Z, Shackley P, Donaldson C, Abdalla M. Counting the benefits of screening: a pilot study of willingness to pay for cystic fibrosis screening. *J Med Screen* 1994;**1**:82–3.
11. Ryan M, Shackley P. Assessing the benefits of health care: how far should we go? *Qual Health Care* 1995;**4**:207–13.
12. Ruwe PA, Wright J, Randall RL, Lynch JK, Jokl P, McCarthy S. Can MR imaging effectively replace diagnostic arthroscopy? *Radiology* 1992;**183**:335–9.
13. Dixon AK. Magnetic resonance imaging of meniscal tears of the knee [editorial]. *J Bone Joint Surg Br* 1996;**78**:174–6.
14. Fischer SP, Fox JM, Del Pizzo W, Friedman MJ, Snyder SJ, Ferkel RD. Accuracy of diagnoses from magnetic resonance imaging of the knee. A multi-centre analysis of one thousand and fourteen patients. *J Bone Joint Surg Am* 1991;**73**:2–10.
15. Lundberg M, Odensten M, Thuomas KA, Messner K. The diagnostic validity of magnetic resonance imaging in acute knee injuries with hemarthrosis. A single-blinded evaluation in 69 patients using high-field MRI before arthroscopy. *Int J Sports Med* 1996;**17**:219–22.
16. Rubin DA, Kettering JM, Towers JD, Britton CA. MR imaging of knees having isolated and combined ligament injuries. *Am J Roentgenol* 1998;**170**:1207–13.
17. Boden SD, Davis DO, Dina TS, Stoller DW, Brown SD, Vailas JC, *et al.* A prospective and blinded investigation of magnetic resonance imaging of the knee. Abnormal findings in asymptomatic subjects. *Clin Orthop* 1992;**282**:177–85.
18. Szczepura AK, Fletcher J, Fitzpatrick JD. Cost effectiveness of magnetic resonance imaging in the neurosciences. *BMJ* 1991;**303**:1435–9.
19. University of York. Economic evaluation of the clinical applications of magnetic resonance imaging. Report to the Medical Research Council. York: University of York; 1990.
20. Chissell HR, Allum RL, Keightley A. MRI of the knee: its cost-effective use in a district general hospital. *Ann R Coll Surg Engl* 1994;**76**:26–9.
21. Barnet MJ. MR diagnosis of internal derangements of the knee: effect of field strength on efficacy. *Am J Roentgenol* 1993;**161**:115–18.
22. James P, Buirski G. MR imaging of the knee: a prospective trial using a low strength magnet. *Australas Radiol* 1990;**34**:59–63.
23. Polly DW, Callaghan JJ, Sikes RA, McCabe JM, McMahon K, Savory CG. The accuracy of selective magnetic resonance imaging compared with the findings of arthroscopy of the knee. *J Bone Joint Surgery Am* 1988;**70**:192–8.
24. Mink JH, Levy T, Crues JV. Tears of the anterior cruciate ligament and menisci of the knee: MR imaging evaluation. *Radiology* 1988;**167**:769–74.
25. Lee JK, Yao L, Phelps CT. Anterior cruciate ligament tears; MR imaging compared with arthroscopy and clinical tests. *Radiology* 1988;**166**:861–4.
26. Crues JV, Mink J, Levy TL, Lotysch M, Stoller DW. Meniscal tears of the knee: accuracy of MR imaging. *Radiology* 1987;**164**:445–8.

27. Glashow JL, Katz R, Schneider M, Scott WN. Double-blind assessment of the value of magnetic resonance imaging in the diagnosis of anterior cruciate and meniscal lesions. *J Bone Joint Surg Am* 1989;**71**:113–19.

28. Warwick DJ, Cavanagh P, Bell M, Marsh CH. Influence of magnetic resonance imaging on a knee arthroscopy waiting list. *Injury* 1993;**24**:380–2.

29. Franken EA, Berbaum KS, Dunn V, Smith WL, Ehrhardt JC, Levitz GS, *et al.* Impact of MR imaging on clinical diagnosis and management: a prospective study. *Radiology* 1986;**161**:377–80.

30. MacKenzie R, Dixon AK, Keene GS, Hollingworth W, Lomas DJ, Villar RN. Magnetic resonance imaging of the knee: assessment of effectiveness. *Clin Radiology* 1996;**51**:245–50.

31. Mackenzie R, Palmer CR, Lomas DJ, Dixon AK. Magnetic resonance imaging of the knee: diagnostic performance statistics. *Clin Radiol* 1996; **51**:251–7.

32. Hollingworth W, Mackenzie R, Todd CJ, Dixon AK. Measuring changes in quality of life following magnetic resonance imaging of the knee: SF-36, EuroQol or Rosser index? *Qual Life Res* 1995;**4**:325–34.

33. Boeree NR, Watkinson AF, Ackroyd CE, Johnson C. Magnetic resonance imaging of meniscal and cruciate injuries of the knee. *J Bone Joint Surg Br* 1991;**73**:452–7.

34. Grevitt MP, Pool CJF, Bodley RN, Savage PE. Magnetic resonance imaging of the knee: initial experience in a district general hospital. *Injury* 1992;**23**:410–12.

35. Jackson DW, Jennings LD, Maywood RM, Berger PE. Magnetic resonance imaging of the knee. *Am J Sports Med* 1988;**16**:29–37.

36. Kelly MA, Flock TJ, Kimmel JA, Kiernan HA, Singson RS, Starron RB, *et al.* MR imaging of the knee: clarification of its role. *Arthroscopy* 1991;**7**:78–85.

37. Rangger C, Klestil T, Kathrein A, Inderster A, Hamid L. Influence of magnetic resonance imaging on indications for arthroscopy of the knee. *Clin Orthop* 1996;**330**:133–42.

38. Trieshmann HW, Mosure JC. The impact of magnetic resonance imaging of the knee on surgical decision making. *Arthroscopy* 1996;**12**:550–5.

39. Spiers ASD, Meagher T, Ostlere SJ, Wilson DJ, Dodd CAF. Can MRI of the knee affect arthroscopic practice? A prospective study of 58 patients. *J Bone Joint Surg Br* 1993;**75**:49–52.

40. Alioto RJ, Browne JE, Barnthouse CD, Scott AR. The influence of MRI on treatment decisions regarding knee injuries. *Am J Knee Surg* 1999;**12**:91–7.

41. Gelb HJ, Glasgow SG, Sapega AA, Torg JS. Magnetic resonance imaging of knee disorders. Clinical value and cost-effectiveness in a sports medicine practice. *Am J Sports Med* 1996;**24**:99–103.

42. Mandelbaum BR, Finerman GAM, Reicher MA, Hartzman S, Bassett LW, Gold RH, *et al.* Magnetic resonance imaging as a tool for evaluation of traumatic knee injuries. Anatomical and pathoanatomical correlations. *Am J Sports Med* 1986;**14**:361–70.

43. Hailey D. An assessment of the status of magnetic resonance imaging in health care. *J Qual Clin Pract* 1997;**17**:221–30.

44. Ware JE, Sherbourne CD. The MOS 36-item Short-Form health survey (SF-36). *Med Care* 1992;**30**:473–83.

45. Brazier JE, Harper R, Jones NMB, O'Cathain A, Thomas KJ, Usherwood T, *et al.* Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *BMJ* 1992;**305**:160–4.

46. Garratt A, Ruta DA, Abdalla MI, Buckingham JK, Russell IT. The SF-36 health survey questionnaire: an outcome measure suitable for routine use within the NHS? *BMJ* 1993;**306**:1440–4.

47. Jenkinson C, Coulter A, Wright L. Short Form-36 (SF-36) health survey questionnaire: normative data for adults of working age. *BMJ* 1993;**306**:1437–40.

48. Brooks RG. How should health status and quality of life be measured? *IHE Information* 1991;**3**:4–5.

49. Hurst NP, Jobanputra P, Hunter M, Lambert M, Lochhead A, Brown H. Validity of EuroQol – a generic health status instrument – in patients with rheumatoid arthritis. *Br J Rheumatol* 1994;**33**:655–62.

50. Brooks R. EuroQol: the current state of play. *Health Policy* 1996;**37**:53–72.

51. Dolan P. Modelling valuations for the EuroQol health states. *Med Care* 1997;**35**:1095–108.

52. Netten A, Dennet J. Unit costs of health and social care. Kent: PSSRU, University of Kent; 1997.

53. Automobile Association. Motoring costs. London: AA Technical Services; 1998.

54. Government Statistical Service. The UK in figures: standard of living. 1998. URL: http://www.statistics.gov.uk/stats/ukinfigs.

55. Bryan S, Brown J, Warren R. Mammography screening: an incremental cost effectiveness analysis of two view versus one view procedures in London. *J Epidemiol Community Health* 1995;**49**:70–8.

56. Brown J, Bryan S, Warren R. Mammography screening: an incremental cost effectiveness analysis of double versus single reading of mammograms. *BMJ* 1996;**312**:809–12.

57. Altman DG. Practical statistics for medical research. London: Chapman and Hall; 1991.

58. Amemiya T. Regression analysis when the dependent variable is truncated. *Normal Econometrica* 1973;**41**:997–1016.

59. Briggs A, Gray A. The distribution of health care costs and their statistical analysis for economic evaluation. *J Health Serv Res Policy* 1998;**3**:233–45.

60. Desgagne A, Castilloux AM, Angers JF, LeLorier J. The use of the bootstrap statistical method for the pharmacoeconomic cost analysis of skewed data. *Pharmacoeconomics* 1998;**13**:487–97.

61. Lawson GM, Nutton RW. A propspective audit of knee arthroscopy: a study of the accuracy of clinical diagnosis and therapeutic value of 325 knee arthroscopies. *J Roy Coll Surg Edinb* 1995;**40**:135–7.

62. Oberlander MA, Shalvoy RM, Hughston JC. The accuracy of the clinical knee examination documented by arthroscopy. A prospective study. *Am J Sports Med* 1993;**21**:773–8.

63. Miller GK. A prospective study comparing the accuracy of the clinical diagnosis of meniscus tear with magnetic resonance imaging and its effect on clinical outcome. *Arthroscopy* 1996;**12**:406–13.

64. Muellner T, Weinstabl R, Schabus R, Vécsei V, Kainberger F. The diagnosis of meniscal tears in athletes. A comparison of clinical and magnetic resonance imaging investigations. *Am J Sports Med* 1997;**25**:7–12.

65. O'Shea KJ, Murphy KP, Heekin RD, Herzwurm PJ. The diagnostic accuracy of history, physical examination, and radiographs in the evaluation of traumatic knee disorders. *Am J Sports Med* 1996;**24**:164–7.

66. Rose NE, Gold SM. A comparison of accuracy between clinical examination and magnetic resonance imaging in the diagnosis of meniscal and anterior cruciate ligament tears. *Arthroscopy* 1996;**12**:398–405.

67. Senghas RE. Indications for magnetic resonance imaging [editorial]. *J Bone Joint Surg Am* 1991;**73**:1.

68. Dandy DJ. Arthroscopy and MRI for the knee [editorial]. *J Bone Joint Surg Br* 1997;**79**:520.

69. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**:159–74.

70. Kundel HL. Medical image perception. *Acad Radiol* 1995;**2**:S108–10.

71. Kelly S, Berry E, Roderick P, Harris KM, Cullingworth J, Gathercole L, *et al.* The identification of bias in studies of the diagnostic performance of imaging modalities. *Br J Radiol* 1997;**70**:1028–35.

72. Machin D, Campbell MJ, Fayers PM. Sample size tables for clinical studies. 2nd ed. Oxford: Blackwell Science; 1997.

73. Fischer SP, Fox JM, Del Pizzo W, Friedman MJ, Snyder SJ, Ferkel RD. Accuracy of diagnoses from magnetic resonance imaging of the knee. A multicenter analysis of one thousand and fourteen patients. *J Bone Joint Surg Am* 1991;**73**:2–10.

74. Raunest J, Oberle K, Loehnert J, Hoetzinger H. The clinical value of magnetic resonance imaging in the evaluation of meniscal disorders. *J Bone Joint Surg Am* 1991;**73**:11–16.

75. Bryan S, Buxton M, Sheldon R, Grant A. Magnetic resonance imaging for the investigation of knee injuries: an investigation of preferences. *Health Econ* 1998;**7**:595–603.

76. Hensher DA, Barnard PO, Truong TP. The role of stated preference methods in studies of travel choice. *J Transport Econ Policy* 1988;**22**:45–58.

77. Mackenzie R, Lomas DJ, Villar RN, Sands DJ. Magnetic resonance imaging for the problem knee: a randomized clinical trial. *Br J Radiol* 1993;**66** (Suppl):21.

78. Birch N, Powles D, Dorrell H, Brooks P. The investigation and treatment of disorders of the knee: indications and a cost-comparison of arthroscopy and magnetic resonance imaging. *Health Trends* 1994;**26**:50–2.

79. Ryan M. Using conjoint analysis to take account of patient preferences and go beyond health outcomes: an application to in vitro fertilisation. *Soc Sci Med* 1999;**48**:535–46.

80. Ryan M. A role for conjoint analysis in technology assessment in health care? *Int J Technol Assess Health Care* 1999;**15**:443–57.

81. Ryan M, Hughes J. Using conjoint analysis to assess women's preferences for miscarriage management. *Health Econ* 1997;**6**:261–73.

82. Bryan S, Parry D. Structural reliability of conjoint measurement in health care: an empirical investigation. *Applied Financial Economics.* In press 2001.

83. Birch N, Powles D, Dorrell H, Brooks P. The investigation and treatment of disorders of the knee: indications and a cost-comparison of arthroscopy and magnetic resonance imaging. *Health Trends* 1994;**26**:50–2.

84. Verbeek M. On the estimation of a fixed effects model. *Economics Letters* 1990;**34**:267–70.

85. Zabel J. Estimating fixed and random effects models with selectivity. *Economics Letters* 1992;**40**:269–72.

86. Wooldridge J. Selection corrections for panel data models under conditional mean independence assumptions. *J Econometrics* 1995;**68**:115–32.

87. Verbeek M, Nijman T. Incomplete panels and selection bias. In: Matyas L, Sevestre P, editors. The econometrics of panel data. Amsterdam: North-Holland; 1996.

88. StataCorp. Stata Statistical Software [computer program]. Release 5.0. College Station, TX: Stata Corporation; 1997.

89. Mundlak Y. On the pooling of time series and cross section data. *Econometrica* 1978;**46**:69–85.

90. Butler J. Moffitt R. A computationally efficient quadrature procedure for the one-factor multinomial probit model. *Econometrica* 1982;**50**:761–4.

91. Lee L. A numerically stable quadrature procedure for the one-factor random-component discrete choice model. *J Econometrics* 2000;**95**:117–29.

92. Lysholm J, Gilquist J. Evaluation of knee ligament surgery results with special emphasis on use of a scoring scale. *Am J Sports Med* 1982;**10**:150–4.

Appendix 1

Patient information sheet and consent form

PATIENT INFORMATION SHEET

INVITATION TO JOIN RESEARCH INTO KNEE INVESTIGATIONS

You are invited to join our research project, which will look at different approaches to the investigation of knee joint problems. At the moment, the best approach for investigating knee joint problems is not known. The objective of this research is to determine the most appropriate approach for patients with knee joint problems. We will look at two alternative approaches. One involves the use of a machine called a magnetic resonance imaging (MRI) scanner to take a picture of the inside of the knee. The other involves a minor operation called arthroscopy to look inside the knee joint.

In this project, half of patients will have an MRI scan as their first investigation. The other half of patients in our project will only have arthroscopy. The decision whether patients receive MRI or arthroscopy will be completely random. It is important to note that about 85% of patients who have an MRI scan will typically need to have an arthroscopy in order to treat their knee problems.

All study patients will be asked to complete a short postal questionnaire asking about their health at the start of the study and at 6 and 12 months later. These will be sent to each patient's home address and will include a stamped addressed envelope for return to the Health Economics Research Group at Brunel University. If you agree to take part in this research, some information about your treatment will be collected from your medical records. **All information collected will be treated as confidential and will not be used in any way that could identify you**.

We would like you to consider taking part in this research. Please complete and return the attached consent form indicating whether or not you wish to take part in the research. If you agree to join and later change your mind, you will still be able to withdraw from the research and have your research record destroyed.

More information about the two methods of investigation is attached. If you would like to discuss this study or have any questions, then the study researcher Mrs Hilary Bungay will be happy to speak to you. She can be contacted by phone at Kent and Canterbury Hospital (01227 766877 ext 4880) on Tuesdays, Wednesdays and Thursdays. If she is unavailable, please leave your name and contact number and she will return your call.

MRI is a relatively new imaging technique for investigating knee problems. The advantages are:

- it is a technique which does not use X-rays
- it is done as an outpatient procedure and takes about 30 minutes
- it **might** avoid the need for an operation: in about 14%[1] of cases MRI is expected to avoid the need for an operation.

The disadvantages are:

- if the MRI scan shows a problem with the knee that requires arthroscopy for treatment, then an extra visit to the clinic may be necessary
- some patients find the examination to be claustrophobic
- MRI is a new technique and has been shown to give misleading information in 8%[2] of knee examinations. In the vast majority of such cases, this has meant that the patient has gone on to have an unnecessary arthroscopy procedure.

Arthroscopy is an established procedure. The advantages are:

- it can be used to diagnose and treat knee disorders at the same time in one operation
- it is a well-accepted procedure and is considered to provide very accurate information on the causes of knee problems
- it is commonly performed as a day-case procedure, even after an MRI scan.

The disadvantages are:

- it requires day-case treatment in hospital under a general anaesthetic
- in less than 5% of cases, there may be a complication, such as a wound infection, and, whilst this is not common, it is serious when it occurs
- it normally takes about 10 days for the knee to recover fully from the operation.

References

1. Birch N, Powles D, Dorrell H, Brooks P. The investigation and treatment of disorders of the knee: indications and a cost-comparison of arthroscopy and magnetic resonance imaging. *Health Trends* 1994;**26**:50–2.

2. Warwick DJ, Cavanagh P, Bell M, Marsh CH. Influence of magnetic resonance imaging on a knee arthroscopy waiting list. *Injury* 1993;**24**:380–2.

PATIENT CONSENT FORM

MRI for investigation of the knee joint

(The patient should complete the whole of this sheet himself/herself) Please cross out as necessary

Have you read the Patient Information Sheet? ..... YES/NO

Have you had an opportunity to ask questions and discuss this study? ..... YES/NO

Have you received satisfactory answers to all of your questions?..... YES/NO

Have you received enough information about the study?..... YES/NO

Dr/Mr..[NAME]..discussed the treatment with me..... YES/NO

Do you understand that you are free to withdraw from the study:

- at any time
- without having to give a reason for withdrawing
- and without affecting your future medical care? ..... YES/NO

Do you agree to take part in this study? ..... YES/NO

Signed ..... Date .....

(NAME IN BLOCK LETTERS): .....

Appendix 2

Sample size calculations

Sample size calculation from the original proposal

"The study will be conducted over 3 years to allow a period of 14 months for patient recruitment. At Kent and Canterbury Hospital, this is expected to allow about 120 patients to be recruited. At St Thomas' Hospital, there remains some uncertainty about the patient number that might be recruited, but a provisional estimate is that it might be possible to generate about 80 patients over the 14-month period. If a total sample of 200 patients were recruited, this would provide a sample with 80% power to detect a statistically significant ($p < 0.05$) difference between the MRI and no-MRI groups on the physical functioning score of the SF-36 of 6.28."

Sample size calculation based on rate of surgery

Let us assume that without MRI the vast majority (i.e. 90%) of patients presenting with a knee injury will be investigated (or directly treated) using arthroscopy. On the basis of published observational data,[83] there is reason to believe that arthroscopy might be avoided in up to 28% of cases by the use of a preliminary MRI investigation. For such a difference between groups (i.e. 90% with no MRI versus 64.8% with MRI) to be established as statistically significant (80% power, $p < 0.05$), the trial would require a total sample of about 100 patients allocated evenly between arms.

Appendix 3

A two-limit tobit model with random effects and sample selectivity

Introduction

Various panel data models that deal simultaneously with the problems of 'unobserved heterogeneity' and 'sample selectivity' have been developed.[84–87] This appendix extends the simultaneous control of unobserved heterogeneity and sample selectivity to the case of continuous variables that are both left- and right-censored. The model can be viewed as an extension of the two-limit censored regression model incorporated in some econometric software packages.[88]

Statistical model

Let i denote the ith individual in a random sample of size N from a population. Observations are made on this sample at an initial time $t = 0$, and at each of T subsequent times, $t = 1,\ldots,T$. The observable variables are:

y_{it} ≡ a limited dependent variable taking any value on the portion of the real line between the individual-specific thresholds $c_{i\min}$ and $c_{i\max}$ inclusive, where $c_{i\min} < c_{i\max}$. This is only observed if the ith individual participates at time t. (Note that the model can accommodate individual-specific thresholds, but this feature is not required for the present study. The dependent health variables in this study are bounded by the same values for all individuals.)
$\boldsymbol{X}_{it}$ ≡ a vector of explanatory variables at time t, corresponding to y_{it}.
z_{it} ≡ a dichotomous (0/1) variable taking the value 1 if the ith individual participates at time t.
$\boldsymbol{W}_{it}$ ≡ a vector of explanatory variables corresponding to z_{it}.

The observed variables y_{it} and z_{it} are related to latent variables y_{it}^* and z_{it}^*, respectively, as follows:

$y_{it} = c_{i\min}$ if $y_{it}^* \leq c_{i\min}$ and $z_{it}^* > 0$
$y_{it} = y_{it}^*$ if $c_{i\min} < y_{it}^* < c_{i\max}$ and $z_{it}^* > 0$
$y_{it} = c_{i\max}$ if $y_{it}^* \geq c_{i\max}$ and $z_{it}^* > 0$
y_{it} is unobserved if $z_{it}^* \leq 0$
$z_{it} = 1$ if $z_{it}^* > 0$, otherwise $z_{it} = 0$.

In turn, y_{it}^* and z_{it}^* are related to $\boldsymbol{X}_{it}$ and $\boldsymbol{W}_{it}$ as follows:

$$y_{it}^* = \boldsymbol{X}_{it}'\beta_t + v_{it} \quad (1)$$
$$v_{it} = \alpha_i + u_{it} \quad (2)$$
$$z_{it}^* = \boldsymbol{W}_{it}'\gamma_t + \varepsilon_{it} \quad (3)$$
$$\varepsilon_{it} = \delta_i + h_{it} \quad (4)$$

The vectors β_t and γ_t are time-varying parameters to be estimated. v_{it} and ε_{it} are composite disturbances consisting of 'random effects' α_i and δ_i and conventional disturbances u_{it} and h_{it}, respectively.

The stochastic assumptions are:

(i) $\alpha_i \sim \text{IN}(0, \sigma_\alpha^2)$ where IN means 'independent normal'
(ii) $\delta_i \sim \text{IN}(0, \sigma_\delta^2)$
(iii) $u_{it} \sim \text{IN}(0, \sigma_u^2)$
(iv) $h_{it} \sim \text{IN}(0, 1)$
(v) $E[u_{is}h_{it}] = \theta\sigma_u$ for $s = t$, $E[u_{is}h_{it}] = 0$ for $s \neq t$, where θ is the correlation parameter of a standardised bivariate normal distribution
(vi) $E[\alpha_i u_{it}] = E[\delta_i u_{it}] = 0$
(vii) $E[\alpha_i h_{it}] = E[\delta_i h_{it}] = 0$
(viii) $E[\alpha_i \delta_i] = 0$
(ix) $E[v_{it}\boldsymbol{X}_{it}] = \boldsymbol{0}$
(x) $E[v_{it}\boldsymbol{W}_{it}] = \boldsymbol{0}$
(xi) $E[\varepsilon_{it}\boldsymbol{X}_{it}] = \boldsymbol{0}$
(xii) $E[\varepsilon_{it}\boldsymbol{W}_{it}] = \boldsymbol{0}$.

Note that the random effects α_i and δ_i are assumed to be uncorrelated with the variables in $\boldsymbol{X}_{it}$ and $\boldsymbol{W}_{it}$ and with each other. If necessary, this may be after imposing structure on an original pair of random effects to eliminate such correlations. For example, following Mundlak[89] and Zabel,[85] random effects that are correlated with explanatory variables can be specified as functions of the means of any time-varying variables, hopefully leaving behind random effects which are largely orthogonal to the regressors.

Define: $\alpha_i^* \equiv \alpha_i/\sigma_\alpha \sim \text{IN}(0, 1)$; $\delta_i^* \equiv \delta_i/\sigma_\delta \sim \text{IN}(0, 1)$; $\rho_1^2 \equiv \sigma_\alpha^2/(\sigma_\alpha^2 + 1)$; $\rho_2^2 \equiv \sigma_\delta^2/(\sigma_\delta^2 + 1)$; $\rho_1^* \equiv \rho_1/(1 - \rho_1^2)^{-1/2}$; $\rho_2^* \equiv \rho_2/(1 - \rho_2^2)^{-1/2}$. Let the sets $R_{i1} = \{c_{i\min}\}$, $R_{i2} = (c_{i\min}, c_{i\max})$ and $R_{i3} = \{c_{i\max}\}$, and let $I(R_{ij})$ be an indicator function

such that $I(R_{ij}) = 1$ if $y_{it} \in R_{ij}$, otherwise $I(R_{ij}) = 0$. Finally, let $\phi(\cdot)$ be the standard normal p.d.f., $\Phi(\cdot)$ the standard normal c.d.f. and $\Phi_2(\cdot, \cdot; \theta)$ the c.d.f. of the standardised bivariate normal distribution with correlation parameter θ. Given the above, the sample likelihood function (to be maximised with respect to β_t and γ_t for $t = 0, 1, \ldots, T$, as well as ρ_1^*, ρ_2^*, θ and σ_u) is:

$$L = \prod_{i=1}^{N} L_i \tag{5}$$

$$= \prod_{i=1}^{N} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ \prod_{t=0}^{T} A^a \times B^b \times C^c \times D^d \right\} \phi(\alpha_i^*) \phi(\delta_i^*) d\alpha_i^* d\delta_i^*$$

where L_i denotes the likelihood for the ith individual, with:

$$A \equiv \Phi\left([c_{i\min} - \boldsymbol{X}_{it}'\beta_t - \rho_1^*\alpha_i^*]/\sigma_u\right) - \Phi_2([c_{i\min} - \boldsymbol{X}_{it}'\beta_t - \rho_1^*\alpha_i^*]/\sigma_u, -\boldsymbol{W}_{it}'\gamma_t - \rho_2^*\delta_i^*; \theta) \tag{6}$$

$$B \equiv \phi\left([y_{it} - \boldsymbol{X}_{it}'\beta_t - \rho_1^*\alpha_i^*]/\sigma_u\right) - \frac{\partial}{\partial y_{it}} \{\Phi_2([y_{it} - \boldsymbol{X}_{it}'\beta_t - \rho_1^*\alpha_i^*]/\sigma_u, - \boldsymbol{W}_{it}'\gamma_t - (\rho_2^*\delta_i^*; \theta)\} \tag{7}$$

$$C \equiv 1 - \Phi\left([c_{i\max} - \boldsymbol{X}_{it}'\beta_t - \rho_1^*\alpha_i^*]/\sigma_u\right) - \Phi(-\boldsymbol{W}_{it}'\gamma_t - \rho_2^*\delta_i^*) + \Phi_2([c_{i\max} - \boldsymbol{X}_{it}'\beta_t - \rho_1^*\alpha_i^*]/\sigma_u, -\boldsymbol{W}_{it}'\gamma_t - \rho_2^*\delta_i^*; \theta) \tag{8}$$

$$D \equiv \Phi(-\boldsymbol{W}_{it}'\gamma_t - \rho_2^*\delta_i^*) \tag{9}$$

and the exponents are $a \equiv I(R_{i1}) \times z_{it}$, $b \equiv I(R_{i2}) \times z_{it}$, $c \equiv I(R_{i3}) \times z_{it}$, $d \equiv 1 - z_{it}$.

To see how this is derived, note that the sample likelihood is:

$$L = \prod_{i=1}^{N} L_i = \prod_{i=1}^{N} P(y_{i0}, z_{i0}, y_{i1}, z_{i1}, \ldots, y_{iT}, z_{iT}) \tag{10}$$

Equation (10) is obtained as equation (5) by considering the conditional density:

$$P(y_{i0}, z_{i0}, y_{i1}, z_{i1}, \ldots, y_{iT}, z_{iT} | \alpha_i^*, \delta_i^*) = \prod_{t=0}^{T} [P(y_{it}, z_{it} = 1 | \alpha_i^*, \delta_i^*)]^z \times [P(z_{it} = 0 | \delta_i^*)]^{1-z} \tag{11}$$

Functional forms for $P(y_{it}, z_{it} = 1 | \alpha_i^*, \delta_i^*)$ and $P(z_{it} = 0 | \delta_i^*)$ in equation (11) are implied by the stochastic assumptions of the model, and it is readily shown that:

$[P(y_{iT}, z_{iT} = 1 | (\alpha_i^*, \delta_i^*)]^z = A^a \times B^b \times C^c$ and
$[P(z_{iT} = 0 | \delta_i^*)]^{1-z} = D^d$

The individual likelihood L_i in equations (5) and (10) is then obtained by 'integrating out' both α_i^* and δ_i^* from the conditional density in equation (11). The integrals in equation (5) can, in theory, be evaluated using Hermite quadrature[90] but this is often difficult in practice, particularly when multiple integrals are involved.[91] A simulation approach was used in this study, based on reconstructing the unobserved random effects in equation (5) by substituting i.i.d. random points for what would have been 'fixed' points in Hermite quadrature. This procedure was implemented using GAUSS 386i for DOS. Full details, including computer programmes, are available from the authors upon request.

Note that θ should lie between 0 and 1 in absolute value, and there are no sample selection effects in the model if $\theta = 0$. Similarly, there are no random effects in the main equations if $\rho_1^* = 0$, and no random effects in the selection equations if $\rho_2^* = 0$. Tests of these null hypotheses constitute tests for the relevance of these effects.

Appendix 4

The cost-effectiveness of MRI in acute knee injuries: the St Thomas' Hospital trial

Introduction

This chapter describes a separate empirical investigation of the use of MRI in the diagnosis of acute knee injuries in a secondary care setting. The principal purpose of this research was to determine whether MRI, when used in patients presenting with an acute knee injury, had a major impact on clinical management, NHS and patient costs and patient outcome. The research was based on a single-centre RCT conducted at St Thomas' Hospital. Research Ethics Committee approval was obtained prior to the commencement of the study.

Methods

Study design and patient recruitment

Patients with acute knee injuries attending the Accident and Emergency Department were referred to a specialist knee clinic. The aim was for trial patients to be representative of the range of knee injuries seen in an Accident and Emergency Department. All patients who attended the knee clinic between April 1996 and July 1997 were assessed for suitability for inclusion in the trial. Patients were defined as suitable for the trial if:

- there had been no previous major surgery in the injured knee, such as knee replacement (previous arthroscopy and partial menisectomy did not exclude patients from the trial)
- there was no pre-existing chronic knee pathology
- there was no serious condition requiring immediate attention, for example, a serious knee infection
- there was no history or current experience of recurrent locking of the knee
- they were aged between 16 and 55 years old.

Once written consent was obtained, study patients were allocated between investigation and treatment informed by an MRI scan (MRI trial arm) or investigation and treatment not informed by MRI (no-MRI trial arm). The on-site researcher undertook the allocation process, following a telephone call to a member of the Trust's Department of Public Health Medicine where randomisation was conducted using a computerised random allocation programme.

The protocol dictated that all study patients, regardless of their trial arm, were referred for an MRI scan (median wait for scan = 2 days). Patients in the MRI arm were reviewed 1 week after the scan by the knee specialist (or by the post-final fellowship senior registrar) and a second diagnosis and management plan was recorded. Only then was the MRI scan reviewed and any change in diagnosis or management plan as a result of this information documented. The clinical follow-up of the patients then continued until their knee recovered. Patients in the no-MRI arm were similarly reviewed 1 week later, but neither the MRI scan nor the report were available in the clinic nor accessible on the hospital computer system. A clinical diagnosis was recorded and a management plan instigated. If a patient in the no-MRI arm still had problems at 6 weeks, the protocol allowed for the MRI scan to be reviewed. Again, any change in management was documented. This protocol allowed assessment of the diagnostic accuracy of clinical examination and MRI separately, as reported in chapter 5.

It was clearly neither feasible nor sensible to blind the study patients, researchers or those involved in providing care to the outcome of the allocation process. At both sites an open-label policy was adopted.

Measurement and analysis

The measurements made and the data analysis methods employed mirrored those used in the trial at Kent and Canterbury Hospital (reported in chapter 4). The only difference was that, at St Thomas' Hospital, patient outcome was additionally measured using a modified version of a knee-specific instrument, Lysholm II.[92] The Lysholm II instrument considers knee-related functioning impairment or problems on eight dimensions: limping, support requirements, locking, instability, pain, swelling, ability to climb stairs and squatting. Each dimension is given a

score, with higher scores reflecting less severe problems, and the dimension scores are summed to give a single overall score (range 0 to 100). The instrument is designed to be administered by a clinician in a patient consultation setting, but was adapted in this study for use in a postal questionnaire survey. As a result, the dimension relating to 'instability' was excluded, since it was not possible to easily convert this into lay terminology. Therefore, the Lysholm II data reported here have a limited possible range of 0 to 75.

Results

Figure 16 shows the trial profile, indicating the number of patients recruited into the study, the randomisation assignment, the numbers receiving an MRI scan and the number of measurements for each randomised group. A total of 120 patients consented to take part in the trial. Of those, 57 were allocated to the MRI arm and 63 to the no-MRI arm. At baseline, the two groups were well matched in terms of age (mean = 28 years) and sex (30% female). The overall response rate to the baseline quality-of-life questionnaire was 94.2% (overall = 113/120, MRI group = 54/57, no-MRI group = 59/63). The baseline comparison of quality-of-life scores is shown in *Figure 17*. Given that data on all SF-36 dimensions and the EQ-5D VAS and tariff had approximately normal distributions, means and 95% CIs are presented. There were no statistically significant differences or evidence of any trends of differences between groups for any of the SF-36 dimensions at baseline (*t*-test minimum $p = 0.24$ for all dimensions). The data indicated that, at the point of recruitment into the trial, patients tended to have particular problems relating to two SF-36 dimensions: bodily pain and role-physical. Similarly, the EQ-5D tariff ($p = 0.79$) and VAS scores ($p = 0.67$) and the

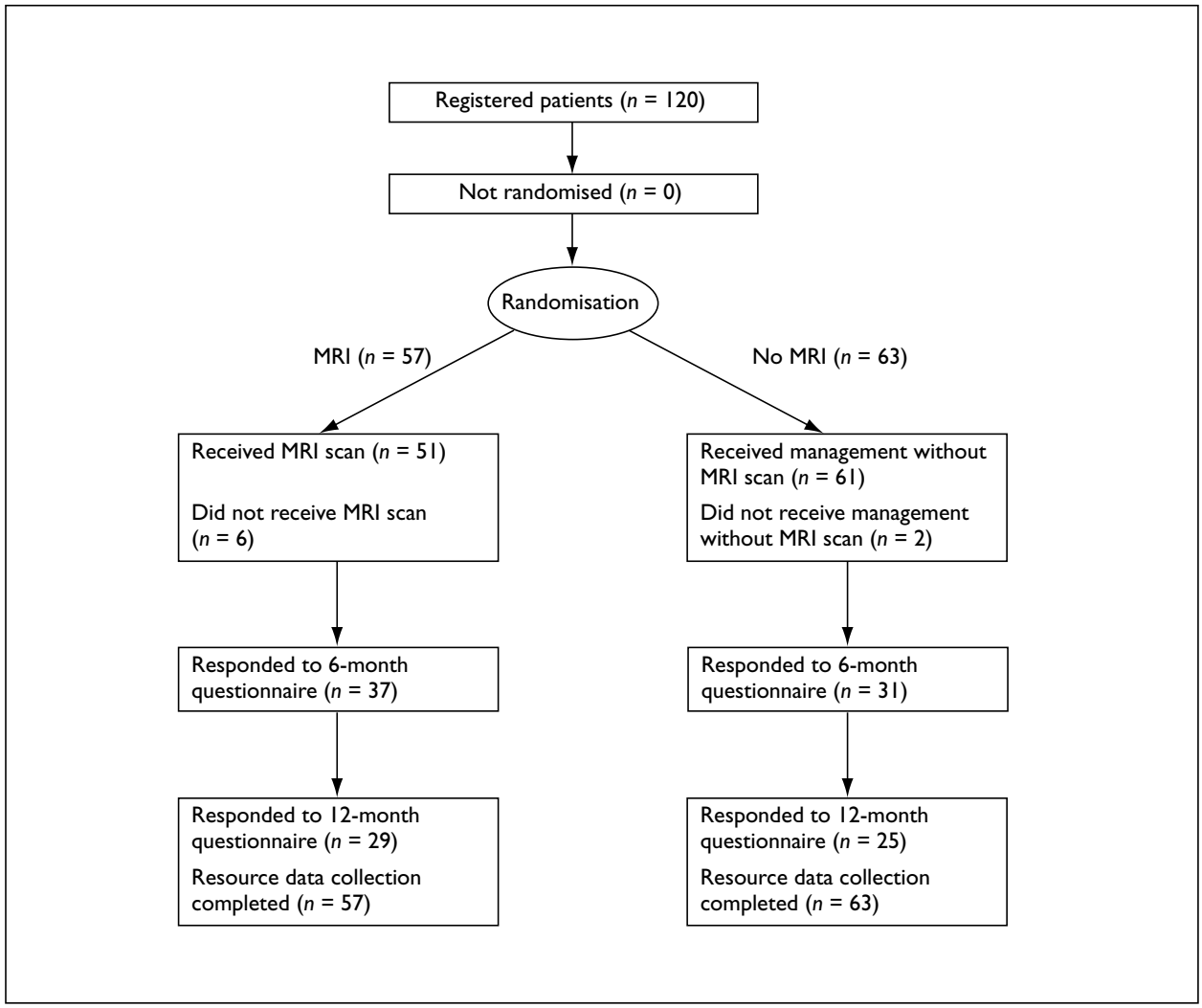


FIGURE 16 *Flow chart describing the progress of patients through the St Thomas' Hospital acute knee trial*

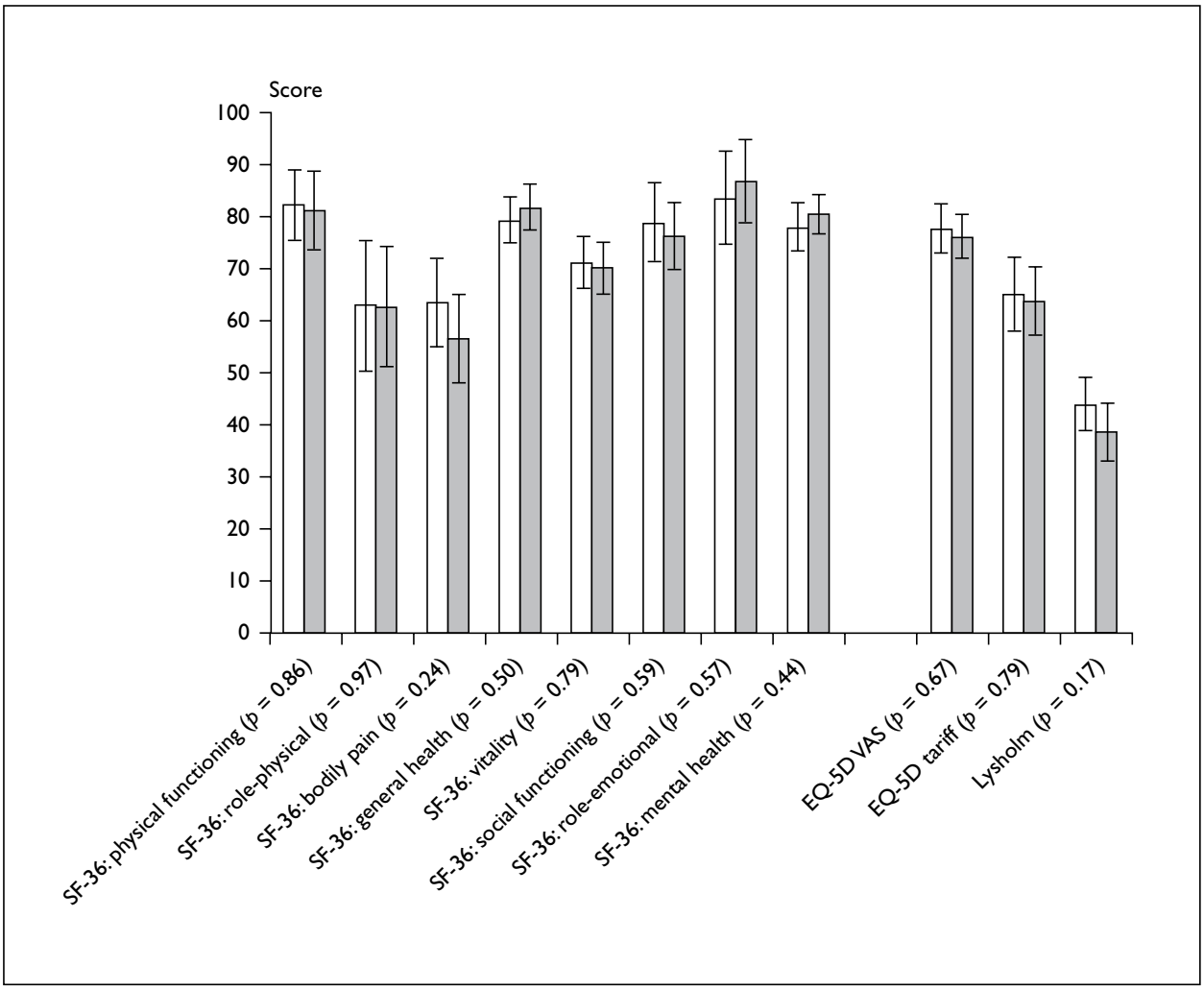


FIGURE 17 *Baseline SF-36 and EQ-5D data of the St Thomas' Hospital acute knee trial (mean scores, 95% CIs) (□, MRI group; ▨, no-MRI group*

Lysholm score ($p = 0.17$) were not statistically significantly different between groups at baseline. The overall response rates to the follow-up quality-of-life questionnaires were 54.1% at 6 months (overall = 68/120, MRI group = 37/57, no-MRI group = 31/63) and 45% at 12 months (overall = 54/120, MRI group = 29/57, no-MRI group = 25/63).

The comparison of quality-of-life scores 6 months after recruitment into the trial are shown in *Figure 18*. Again, the data on all SF-36 dimensions and the EQ-5D VAS and tariff had an approximately normal distribution and thus means and 95% CIs are presented. There were no statistically significant differences between groups for any of the SF-36 dimensions at 6 months after recruitment (t-test minimum $p = 0.50$ for all dimensions), and the data suggested that, for many patients in both groups, problems existed principally in terms of vitality. Similarly, the EQ-5D tariff and VAS scores and the Lysholm scores were not significantly different between groups at 6 months (t-test $p = 0.96$, $p = 0.53$ and $p = 0.96$, respectively).

The general pattern in the quality-of-life data seen at 6 months was repeated in the follow-up data at 12 months. Data on means and CIs are presented in *Figure 19* (approximately normal distributions for all dimensions) and indicate a possible trend of between-group differences: the mean scores for the no-MRI group are higher across most dimensions.

Information on NHS resource use and costs, broken down by trial arm, is presented in *Table 27*. A similar pattern of resource use was found in the two groups in terms of the proportion of patients who received surgery. *Table 27* also reports estimates of the NHS cost per patient. A statistically significant difference was found between mean costs ($p < 0.05$), which indicated that the mean NHS cost per patient was higher, overall, for

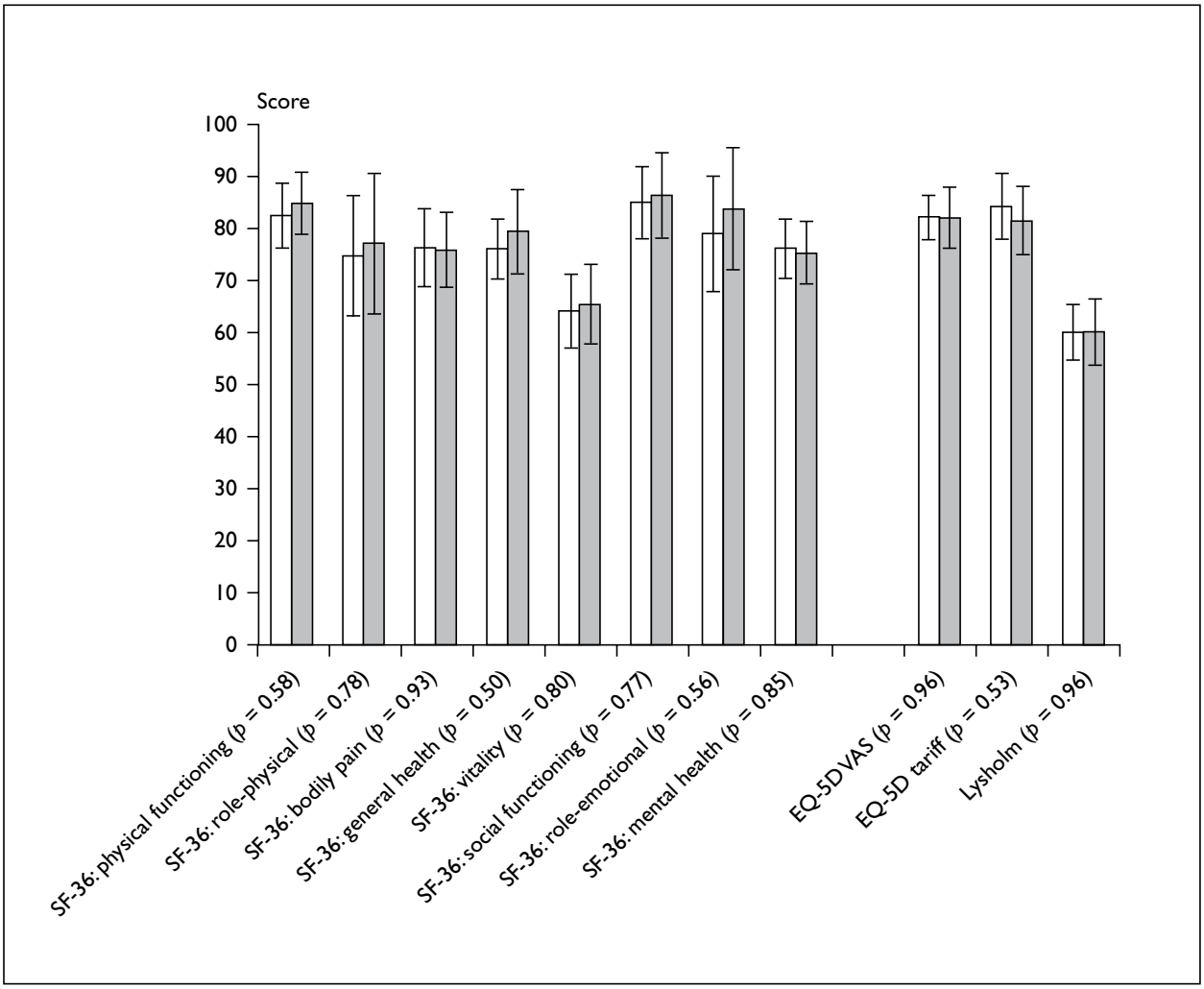


FIGURE 18 *SF-36 and EQ-5D 6-month data of the St Thomas' Hospital acute knee trial (mean scores, 95% CIs) (□, MRI group; ▨, no-MRI group*

patient management with MRI. In addition, *Table 27* shows results for patient costs and for both NHS and patient costs combined. The mean patient cost was again higher in the MRI group, given that virtually all patients in that group had the additional travel and time costs associated with attending for the MRI scan. When all costs were considered (both NHS and patient costs) the results mirror those for NHS costs only: a statistically significant difference between means overall.

Summary of main findings

In terms of patient health outcomes, judged from a traditional health economics perspective, the central finding was that there was no significant difference between groups in health-related utility (measured using the EQ-5D tariff scores). Similarly, on all other measures of health outcome (i.e. EQ-5D VAS and all SF-36 dimensions) no statistically significant differences between groups were revealed. In addition, the use of MRI in this group of patients with acute knee injuries was not found to be associated with a positive diagnostic and therapeutic impact: a similar proportion of patients in the MRI group underwent surgery in the 12-month follow-up period (30% of MRI patients compared to 24% of no-MRI patients). In terms of costs to the NHS, the data suggested a significantly lower mean cost for the no-MRI group. On average, the additional cost associated with providing MRI scans to all patients with acute knee injuries was not offset by any reduction in the proportion of patients who underwent surgery.

Therefore, the study identified a dominated technology: the use of MRI was found to be associated with an increased cost without any improvement in health outcomes or diagnostic impact.

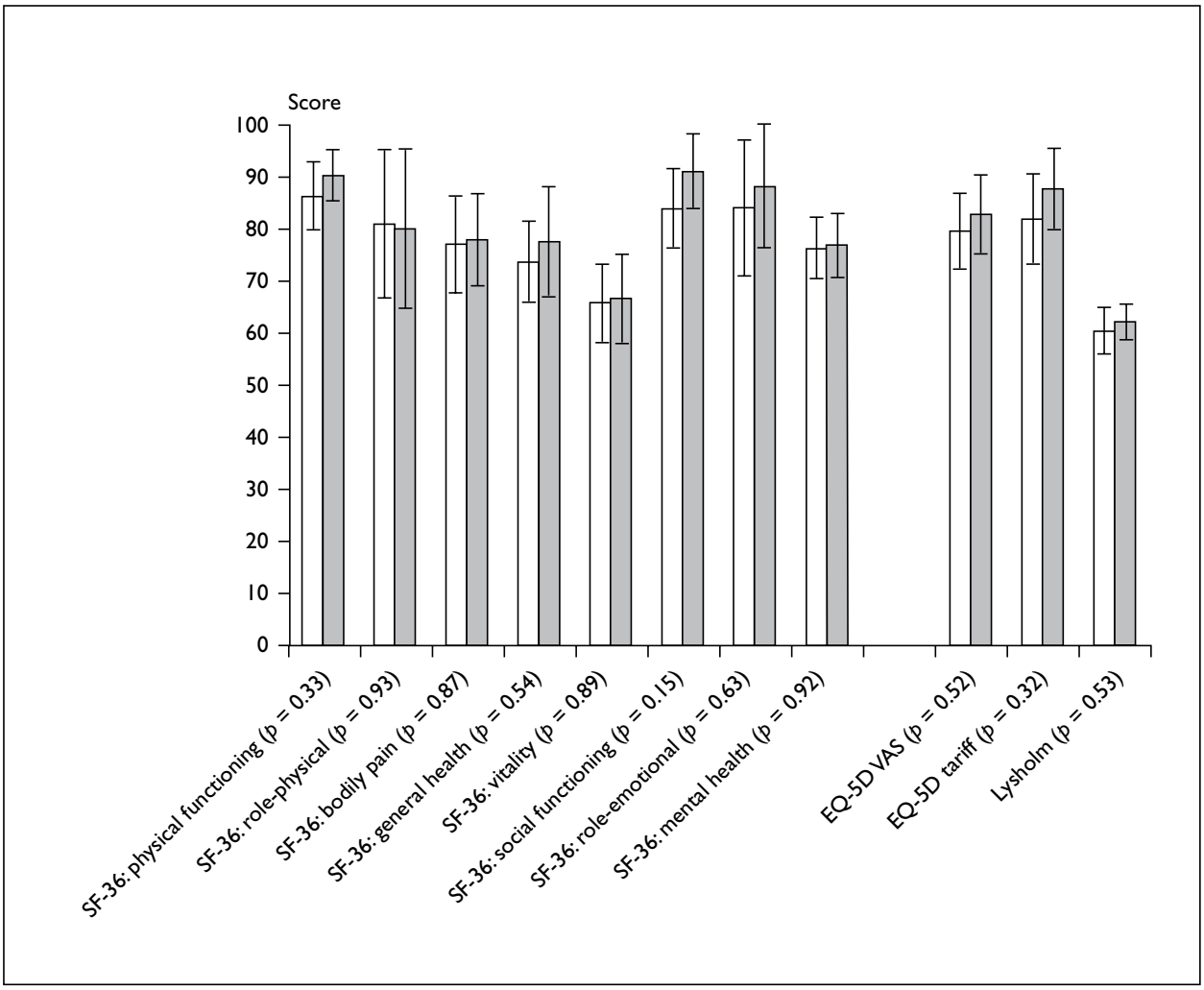


FIGURE 19 *SF-36 and EQ-5D 12-month data of the St Thomas' Hospital acute knee trial (mean scores, 95% CIs) (□, MRI group; ▨, no-MRI group*

TABLE 27 *NHS resource use and costs (£) over 12 months (St Thomas' Hospital acute knee trial)*

	MRI (*n* = 57)	**MRI 95% CI**	**No-MRI (*n* = 63)**	**No-MRI 95% CI**
Proportion undergoing surgery	0.30	0.18 to 0.44	0.24	0.14 to 0.37
Mean (SD) total NHS costs[a]	527 (250)	465 to 599	318 (228)	264 to 376
Mean (SD) total patient costs[a]	138 (31)	131 to 147	108 (33)	100 to 116
Mean (SD) total NHS + patient costs[a]	666 (266)	589 to 730	425 (244)	362 to 483

[a] *Bootstrap comparison of means, 95% CI (bias corrected and accelerated method, 2000 replications)*

Appendix 5

Pro-forma used in image interpretation study

MRI study of the knee

Viewer number ☐ Study number ☐

Please complete all sections and tick one box in each row

A	Time at start of viewing	☐	(24-hour clock)			
B	Knee reported	Right	Left			
		☐ 1	☐ 2			
C	Medial meniscus	Not seen	Normal	Tear	Degenerative changes	
	- Anterior third	☐ 1	☐ 2	☐ 3	☐ 4	
	- Body	☐ 1	☐ 2	☐ 3	☐ 4	
	- Posterior third	☐ 1	☐ 2	☐ 3	☐ 4	
D	Lateral meniscus	Not seen	Normal	Tear	Degenerative changes	
	- Anterior third	☐ 1	☐ 2	☐ 3	☐ 4	
	- Body	☐ 1	☐ 2	☐ 3	☐ 4	
	- Posterior third	☐ 1	☐ 2	☐ 3	☐ 4	
E	Cruciate ligament	Not seen	Normal	Partial tear	Complete rupture	
	- Anterior	☐ 1	☐ 2	☐ 3	☐ 4	
	- Posterior	☐ 1	☐ 2	☐ 3	☐ 4	
F	Collateral ligaments	Not seen	Normal	Sprain	Partial tear	Complete tear
	- Medial	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
	- Lateral	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
G	Bony surfaces	Normal	Fracture	Bone bruise	Degenerative changes	
	- Medial condyle	☐ 1	☐ 2	☐ 3	☐ 4	
	- Lateral condyle	☐ 1	☐ 2	☐ 3	☐ 4	
	- Medial plateau	☐ 1	☐ 2	☐ 3	☐ 4	
	- Lateral plateau	☐ 1	☐ 2	☐ 3	☐ 4	
H	Joint effusion	Not present	Present	Small	Large	
		☐ 1	☐ 2	☐ 3	☐ 4	
J	Time at end of viewing	☐	(24-hour clock)			

K Are you happy with the quality of the images? Yes ☐ 1 No ☐ 2

Thank you very much for your help with this study

Appendix 6

Instructions to readers in image interpretation study

MRI KNEE STUDY

Please follow these instructions for the viewing and scoring of the knee images.

Order of viewing
View the cases in any order. Note that the numbers are not consecutive.

Location of viewing
View the images where you would **normally** report MRI knee studies.

Recording your interpretation of the images

- Each film packet contains the MRI images of one patient and the request form details.
- Forms are provided on which your viewer number has already been recorded. There is one form for each patient.
- In the box on the form called 'study number', please insert the study number that is on the white label on the front of the film packet.
- All forms are held together by a treasury tag so that we know the order in which the films have been viewed.
- **Please follow your normal procedure for reporting MRI knee studies.** If you normally refer to a book or a colleague, please do so.

Completed forms
Please send completed forms to me in the enclosed envelope.

MRI images
Please send these on to the next viewer by courier or Royal Mail, whichever is cheaper for heavy items (the post room will be able to advise you), and send a claim to me for payment. Address labels for the next viewer are enclosed. You may find it useful to reuse the box in which the films arrived.

Payment for undertaking the study
I will send you a claim form on receipt of your completed forms.

If you have any queries about this study, please contact me.

Thank you very much for your help.

Gwyn Weatherburn
HERG
Brunel University
Uxbridge
Middlesex
UB8 3PH

Telephone 01895 274000 ext 3477

email: gwyneth.weatherburn@brunel.ac.uk

Appendix 7

Information sheet and example choice used in conjoint study

INFORMATION SHEET

Knee injuries can be both painful and disabling.

Option 1: Conventional approach to diagnosis and treatment of knee injuries
If the knee injury is severe then 'keyhole' surgery, called arthroscopy, is commonly used to diagnose the problem.

- Arthroscopy is usually done as a day-case under general anaesthetic.
- If the surgeon discovers a problem that requires surgical repair then this can be done at the same time. A second operation is not required.
- In less than 5% of cases, there may be a complication, such as a wound infection.
- Recovery takes between 4 and 6 weeks, during which time the knee will be swollen and mildly painful. Normal activities, including sports, can then be resumed.
- The knee may not require surgical repair and the arthroscopy will have been used to diagnose, but not to treat the problem. In such cases, treatment will involve physiotherapy for a period of 6 weeks.
- Arthroscopy is available free on the NHS.

Option 2: A new approach to diagnosis and treatment of knee injuries
A new method for diagnosing knee problems has been developed, called magnetic resonance imaging (MRI).

- The MRI scan is an outpatient procedure and takes approximately 30 minutes.
- The patient has to place his/her knee inside a large cylinder that contains a powerful magnet.
- There are no known side-effects from an MRI scan of the knee.
- The MRI scan may show that an arthroscopy is necessary to treat the problem.
- If surgery is not required, treatment will involve a 6-week course of physiotherapy.
- MRI is not always available free on the NHS. Patients may have to pay for it privately.

Treatment effectiveness
In many cases, treatment is 100% successful and individuals can resume normal activities with no pain. However, in some cases, treatment is not successful and the injured knee will continue to be painful, especially after vigorous sporting activities, such as rugby and football.

Waiting times
Waiting times for diagnosis and treatment of knee injuries vary enormously from hospital to hospital. The waiting time for an arthroscopy in one hospital may be shorter than the waiting time for physiotherapy in another.

EXAMPLE CHOICE

	Conventional treatment	MRI
Cost to you personally	£0	£400
Chance of avoiding surgery	0%	40%
Time from initial visit to end of treatment	12 weeks	24 weeks
Chance of the knee problem being completely resolved	80%	70%
	☐	☐

Health Technology Assessment Programme

Prioritisation Strategy Group

Members

Chair
Professor Kent Woods
Director,
NHS HTA Programme, &
Professor of Therapeutics
University of Leicester

Professor Bruce Campbell
Consultant General Surgeon
Royal Devon & Exeter Hospital

Professor Shah Ebrahim
Professor of Epidemiology
of Ageing
University of Bristol

Dr John Reynolds
Clinical Director
Acute General Medicine SDU
Oxford Radcliffe Hospital

Dr Ron Zimmern
Director, Public Health
Genetics Unit
Strangeways Research
Laboratories, Cambridge

HTA Commissioning Board

Members

Programme Director
Professor Kent Woods
Director, NHS HTA
Programme, &
Professor of Therapeutics
University of Leicester

Chair
Professor Shah Ebrahim
Professor of Epidemiology
of Ageing
University of Bristol

Deputy Chair
Professor Jon Nicholl
Director, Medical Care
Research Unit
University of Sheffield

Professor Douglas Altman
Director, ICRF Medical
Statistics Group
University of Oxford

Professor John Bond
Director, Centre for Health
Services Research
University of Newcastle-
upon-Tyne

Ms Christine Clark
Freelance Medical Writer
Bury, Lancs

Professor Martin Eccles
Professor of
Clinical Effectiveness
University of Newcastle-
upon-Tyne

Dr Andrew Farmer
General Practitioner &
NHS R&D
Clinical Scientist
Institute of Health Sciences
University of Oxford

Professor Adrian Grant
Director, Health Services
Research Unit
University of Aberdeen

Dr Alastair Gray
Director, Health Economics
Research Centre
Institute of Health Sciences
University of Oxford

Professor Mark Haggard
Director, MRC Institute
of Hearing Research
University of Nottingham

Professor Jenny Hewison
Senior Lecturer
School of Psychology
University of Leeds

Professor Alison Kitson
Director, Royal College of
Nursing Institute, London

Dr Donna Lamping
Head, Health Services
Research Unit
London School of Hygiene
& Tropical Medicine

Professor David Neal
Professor of Surgery
University of Newcastle-
upon-Tyne

Professor Gillian Parker
Nuffield Professor of
Community Care
University of Leicester

Dr Tim Peters
Reader in Medical Statistics
University of Bristol

Professor Martin Severs
Professor in Elderly
Health Care
University of Portsmouth

Dr Sarah Stewart-Brown
Director, Health Services
Research Unit
University of Oxford

Professor Ala Szczepura
Director, Centre for Health
Services Studies
University of Warwick

Dr Gillian Vivian
Consultant in Nuclear
Medicine & Radiology
Royal Cornwall Hospitals Trust
Truro

Professor Graham Watt
Department of
General Practice
University of Glasgow

Dr Jeremy Wyatt
Senior Fellow
Health Knowledge
Management Centre
University College London

continued

continued

Diagnostic Technologies & Screening Panel

Members

Chair
Dr Ron Zimmern
Director, Public Health Genetics Unit
Strangeways Research Laboratories
Cambridge

Dr Philip J Ayres
Consultant in Epidemiology & Public Health
The Leeds Teaching Hospitals NHS Trust

Mrs Stella Burnside
Chief Executive, Altnagelvin Hospitals Health & Social Services Trust
Londonderry
Northern Ireland

Dr Paul O Collinson
Consultant Chemical Pathologist & Senior Lecturer
St George's Hospital, London

Dr Barry Cookson
Director, Laboratory of Hospital Infection
Public Health Laboratory Service, London

Professor Howard Cuckle
Professor of Reproductive Epidemiology
University of Leeds

Dr Carol Dezateux
Senior Lecturer in Paediatric Epidemiology
Institute of Child Health
London

Professor Adrian K Dixon
Professor of Radiology
Addenbrooke's Hospital
Cambridge

Mr Steve Ebdon-Jackson
Head, Diagnostic Imaging & Radiation Protection Team
Department of Health, London

Dr Tom Fahey
Senior Lecturer in General Practice
University of Bristol

Dr Andrew Farmer
General Practitioner & NHS Clinical Scientist
Institute of Health Sciences
University of Oxford

Mrs Gillian Fletcher
Antenatal Teacher & Tutor
National Childbirth Trust
Reigate

Professor Jane Franklyn
Professor of Medicine
University of Birmingham

Dr JA Muir Gray
Joint Director, National Screening Committee
NHS Executive, Oxford

Dr Peter Howlett
Executive Director – Development
Portsmouth Hospitals NHS Trust

Professor Alistair McGuire
Professor of Health Economics
City University, London

Mrs Kathlyn Slack
Professional Support
Diagnostic Imaging & Radiation Protection Team
Department of Health
London

Mr Tony Tester
Chief Officer, South Bedfordshire Community Health Council
Luton

Pharmaceuticals Panel

Members

Chair
Dr John Reynolds
Clinical Director – Acute General Medicine SDU
Oxford Radcliffe Hospital

Dr Felicity J Gabbay
Managing Director, Transcrip Ltd
Milford-on-Sea, Hants

Mr Peter Golightly
Director, Trent Drug Information Services
Leicester Royal Infirmary

Dr Alastair Gray
Director, Health Economics Research Centre
Institute of Health Sciences
University of Oxford

Mrs Jeannette Howe
Senior Principal Pharmacist
Department of Health, London

Dr Andrew Mortimore
Consultant in Public Health Medicine
Southampton & South West Hants Health Authority

Mr Nigel Offen
Head of Clinical Quality
NHS Executive – Eastern
Milton Keynes

Professor Robert Peveler
Professor of Liaison Psychiatry
Royal South Hants Hospital
Southampton

Mrs Marianne Rigge
Director, College of Health
London

Dr Frances Rotblat
Manager, Biotechnology Group
Medicines Control Agency
London

Mr Bill Sang
Chief Executive
Salford Royal Hospitals NHS Trust

Dr Eamonn Sheridan
Consultant in Clinical Genetics
St James's University Hospital
Leeds

Mrs Katrina Simister
New Products Manager
National Prescribing Centre
Liverpool

Dr Ross Taylor
Senior Lecturer
Department of General Practice & Primary Care
University of Aberdeen

Dr Richard Tiner
Medical Director
Association of the British Pharmaceutical Industry
London

Professor Jenifer Wilson-Barnett
Head, Florence Nightingale Division of Nursing & Midwifery
King's College, London

Mr David J Wright
Chief Executive
International Glaucoma Association, London

Therapeutic Procedures Panel

Members

Chair
Professor Bruce Campbell
Consultant General Surgeon
Royal Devon & Exeter Hospital

Professor John Bond
Professor of
Health Services Research
University of Newcastle-upon-Tyne

Ms Judith Brodie
Head of Cancer
Support Service
Cancer BACUP, London

Ms Tracy Bury
Head of Research
& Development
Chartered Society of
Physiotherapy, London

Mr Michael Clancy
Consultant in A&E Medicine
Southampton General Hospital

Professor Collette Clifford
Professor of Nursing
University of Birmingham

Dr Katherine Darton
Information Unit
MIND – The Mental Health
Charity, London

Mr John Dunning
Consultant Cardiothoracic
Surgeon
Papworth Hospital NHS Trust
Cambridge

Mr Jonothan Earnshaw
Consultant Vascular Surgeon
Gloucestershire Royal Hospital

Professor David Field
Professor of Neonatal Medicine
The Leicester Royal Infirmary
NHS Trust

Professor FD Richard Hobbs
Professor of Primary Care
& General Practice
University of Birmingham

Mr Richard Johanson
Consultant & Senior Lecturer
North Staffordshire Infirmary
NHS Trust, Stoke-on-Trent

Dr Duncan Keeley
General Practitioner
Thame, Oxon

Dr Phillip Leech
Principal Medical Officer
Department of Health, London

Professor James Lindesay
Professor of Psychiatry for
the Elderly
University of Leicester

Professor Rajan Madhok
Director of Health Policy
& Public Health
East Riding & Hull
Health Authority

Dr Mike McGovern
Branch Head
Department of Health
London

Dr John C Pounsford
Consultant Physician
Frenchay Healthcare Trust
Bristol

Dr Mark Sculpher
Senior Research Fellow in
Health Economics
University of York

Dr Ken Stein
Consultant in Public
Health Medicine
North & East Devon
Health Authority, Exeter

Expert Advisory Network

Members

Professor John Brazier
Director of Health Economics
University of Sheffield

Mr Shaun Brogan
Chief Executive, Ridgeway
Primary Care Group
Aylesbury, Bucks

Mr John A Cairns
Director, Health Economics
Research Unit
University of Aberdeen

Dr Nicky Cullum
Reader in Health Studies
University of York

Professor Pam Enderby
Chair of Community
Rehabilitation
University of Sheffield

Mr Leonard R Fenwick
Chief Executive
Freeman Hospital
Newcastle-upon-Tyne

Ms Grace Gibbs
Deputy Chief Executive
West Middlesex
University Hospital

Dr Neville Goodman
Consultant Anaesthetist
Southmead Hospital, Bristol

Professor Robert E Hawkins
CRC Professor & Director of
Medical Oncology
Christie Hospital NHS Trust
Manchester

Professor Allen Hutchinson
Director of Public Health &
Deputy Dean, ScHARR
University of Sheffield

Professor David Mant
Professor of General Practice
Institute of Health Sciences
University of Oxford

Professor Alexander Markham
Director
Molecular Medicine Unit
St James's University Hospital
Leeds

Dr Chris McCall
General Practitioner
Corfe Mullen, Dorset

Dr Peter Moore
Freelance Science Writer
Ashtead, Surrey

Dr Sue Moss
Associate Director, Cancer
Screening Evaluation Unit
Institute of Cancer Research
Sutton, Surrey

Mrs Julietta Patnick
National Coordinator
NHS Cancer Screening
Programmes, Sheffield

Professor Jennie Popay
Professor of Sociology &
Community Health
University of Salford

Professor Chris Price
Professor of
Clinical Biochemistry
St Bartholomew's & The
Royal London School of
Medicine & Dentistry

Mr Simon Robbins
Chief Executive
Camden & Islington
Health Authority, London

Dr William Rosenberg
Senior Lecturer &
Consultant in Medicine
University of Southampton

Dr Sarah Stewart-Brown
Director, Health Services
Research Unit
University of Oxford

Dr Gillian Vivian
Consultant in Nuclear
Medicine & Radiology
Royal Cornwall Hospitals Trust
Truro

Mrs Joan Webster
Former Chair
Southern Derbyshire
Community Health Council
Nottingham

Feedback

The HTA Programme and the authors would like to know your views about this report.

The Correspondence Page on the HTA website (http://www.ncchta.org) is a convenient way to publish your comments. If you prefer, you can send your comments to the address below, telling us whether you would like us to transfer them to the website.

We look forward to hearing from you.

Copies of this report can be obtained from:

The National Coordinating Centre for Health Technology Assessment,
Mailpoint 728, Boldrewood,
University of Southampton,
Southampton, SO16 7PX, UK.
Fax: +44 (0) 23 8059 5639 Email: hta@soton.ac.uk
http://www.ncchta.org

ISSN 1366-5278