The problem in a clinical trial for a disease modifying osteoarthritis drug (DMOAD) is measuring very small changes in a disease that progress very slowly. Unless a measurement is both sensitive and highly reproducible, a very large patient cohort must be followed for a number of years. This can make OA trials very expensive and therefore commercially risky.

The usual endpoint of an OA trial is joint space narrowing – or a change in minimum joint space width (mJSW) – measured on a radiograph of the knee, which is a surrogate for cartilage loss. However, modern imaging approaches recognize that OA is a whole joint disease which may involve multiple tissues. In addition, the inability of radiographs to visualize cartilage results in a lack of sensitivity to early and small changes in this tissue. There is also difficulty in obtaining high quality reproducible images of OA joints, despite state-of-the-art standardisation of radiographic protocols to reduce the variability related to joint repositioning[1].

Introduction

Due to the limitations of radiography, MRI has been identified by the Outcome Measures in Rheumatologic Clinical Trials (OMERACT) and OARSI as the most appropriate imaging modality to assess joint status in OA research studies[2]. MRI can detect structural pathology associated with pain and other tissues involved in the disease process, including:

- Bone surface area and shape
- Cartilage damage
- Osteophytes
- Subchondral cysts
- Joint effusions
- Ligament and tendon tears
- Baker’s cysts
- Synovitis
- Meniscal tears
- Subchondral bone marrow lesions

Our work in osteoarthritis has involved the careful measurement of bone, articular cartilage and menisci. Imorphics technology provides very precise measurements of established tissue change such as the loss of articular cartilage, but in addition, we have shown that the bone itself is a highly responsive tissue, usually ignored in clinical trials. Imorphics has pioneered the use of 3D bone shape as a sensitive and repeatable biomarker of OA, reducing commercial risk with early go/no-go decisions. Our work is facilitating the delivery of osteoarthritis trials of the future with smaller patient cohorts and shorter time scales, providing new insights into disease progression and making trials viable that might otherwise have been uneconomic.

The Use of MRI

The Use of MRI

At left: Automated segmentation of an osteoarthritic knee
This knee shows advanced osteophytes. As well as the major bones, articular cartilage and the meniscus are included.
involved fully automated segmentation of knee bone and cartilage from MR images. Our delivered software performed fully automated segmentation of the femur, tibia, femoral cartilage and tibial cartilage with no additional manual correction. Our average distance error ranked first out of 19 groups, with an excellent $0.42 \pm 0.74$ mm for the femur and $0.38 \pm 0.68$ mm for the tibia. In cartilage volume difference scores, we again ranked first with an average volume error of 4.2%.

Importantly, our latest technical developments mean that we can now segment the knee with average distance errors of around 0.1 mm. It is this accurate and automated - and therefore, precise – segmentation technology that forms the basis of our OA clinical trials methodology.

In particular, the bone surface and automated landmark placement that comes from using statistical models provides a frame of reference for measurement comparison of within or between subject change and variation.

**Imorphics Automated Segmentation Technology**

Imorphics fully-automated identification and segmentation solutions are usually specified with sub-voxel or sub-millimetre accuracy. Reproducibility is excellent, with typical CoVs of around 1%. After segmentation, the object surface is described by a dense set of true landmarks that correspond.

These can be used to define very accurate measurements between points or areas of interest as patches. In addition, statistical shape model methods can readily deal with cropped images or missing anatomy.

Each year, the prestigious Medical Image Computing and Computer Assisted Intervention Society (MICCAI) holds a "Grand Challenge"[3]. This competition allows entrants from both academia and industry to test their methods in a direct comparison with the state-of-the-art on previously unseen medical images.

The testing is done live and concurrently in order to give a fair representation of clinical performance. Imorphics has won all four of the MICCAI Grand Challenge segmentation competitions that we have entered.

In 2010, Imorphics took part in the “Segmentation of Knee Images” (SKI10)[4] competition which involved fully automated segmentation of knee bone and cartilage from MR images. Our delivered software performed fully automated segmentation of the femur, tibia, femoral cartilage and tibial cartilage with no additional manual correction. Our average distance error ranked first out of 19 groups, with an excellent $0.42 \pm 0.74$ mm for the femur and $0.38 \pm 0.68$ mm for the tibia. In cartilage volume difference scores, we again ranked first with an average volume error of 4.2%.

Importantly, our latest technical developments mean that we can now segment the knee with average distance errors of around 0.1 mm. It is this accurate and automated - and therefore, precise – segmentation technology that forms the basis of our OA clinical trials methodology.

In particular, the bone surface and automated landmark placement that comes from using statistical models provides a frame of reference for measurement comparison of within or between subject change and variation.

**Anatomically Corresponded Regional Analysis of Cartilage (ACRAC)**

In MRI studies of knee OA, cartilage thickness and volume are commonly investigated as morphological parameters. Cross-sectional studies have shown that cartilage of normal controls is thicker than cartilage of patients with knee OA, while longitudinal studies have shown cartilage thinning with the progression of the pathology[5].
Automated Cartilage Segmentation

More recently we have compared an updated version of our fully automated cartilage segmentation method[4] with careful manual segmentation[10]. In a study of 565 knees from the Osteoarthritis Initiative over a 2-year period, we compared the responsiveness of cartilage thickness in the central medial femur region (cMF) using either automatic segmentation or careful manual segmentation. We also compared the agreement between the two methods.

For automated cartilage segmentation, each image is automatically segmented using AAMs of bone and cartilage through multi-start optimisation. Initially, this fits low-density low resolution models but ends in a robust matching of detailed high resolution models. Finally, the voxels contained in the cartilage region are assigned using a probabilistic model learnt during training.

We found that the methods agreed well, with a systematic bias of -0.034mm, with a 95% confidence limit of 0.37mm, comparable to our own manual test-retest agreement. The change in manual cartilage thickness (cMF.ThCtAB) at 1 year was 0.037mm, 95% confidence limit (0.028, 0.046) and at 2 years was 0.059mm (0.047, 0.081). The change in automated cMF at 1 years was 0.061mm (0.048, 0.074), at 2 years was 0.090mm (0.075, 0.105), showing improved sensitivity over the manual method (see Figure2).

The study showed that automated cartilage segmentation using AAMs now provides comparable cartilage thickness measures to careful manual segmentation, but with improved responsiveness.

Because manual cartilage segmentation is very labour-intensive, it can limit the use of measurements derived from MRI in OA clinical trials, however, automation now provides an equally accurate alternative, enabling the economically viable segmentation of large datasets.

Figure 2: Change in cartilage thickness for automated (Imorphics cMF) and manual segmentations (cMF.ThCtAB) SRM values are shown alongside the relevant points.
Measurement of Bone Shape and Progression of OA

Because knee OA is thought to be a largely mechanically driven process[11], a promising target for an OA imaging biomarker may be to exploit the ability of bone to adapt to mechanical influences[12]. In particular, bone can readily change its shape in response to stresses acting upon it (Wolff’s law), suggesting that such alterations may be feasibly assessed in a practical timeframe, making it attractive as a potential imaging end point for trials. Additionally, subtle differences in bone shape or geometry itself could lead to abnormal joint loading and a predisposition to OA.

This method involves the automatic segmentation of MR images using 3D active appearance models, which are a type of shape model designed for searching inside images[6]. Using a statistical shape model in this way allows for the full parameterisation of the shape of each subject knee in terms of the population mean and shape variation learnt during the model training phase.

This parameterisation can then be used to construct a vector of OA versus non-OA shape. This vector, trained on OA and non-OA shapes can identify knees without OA that are at risk of developing OA 12 months later and beyond, and that the position along this vector is associated with OA incidence[13].

Measurement of Changes in Bone Area

It is recognized that OA is a whole joint disease that may involve multiple tissues which confer different phenotypes[14], and subchondral bone is integral to the pathogenesis and progression of OA[14],[15]. The area of subchondral bone at the femorotibial articulation is larger in OA knees than healthy controls, and correlates with knee joint space narrowing, osteophytes and Kellgren Lawrence (KL) grade after adjusting for appropriate confounders in cross-sectional studies.

Radiographic measures, derived from a single radiographic projection, are only weakly associated with OA-attributable bone area measured in 3D. This may reflect the additional 3D MRI structural information, unaccounted for by these 2D radiographic measures[16],[17].

The use of 3D statistical shape models means that each bone surface is fitted with a dense set of landmarks during auto-segmentation. The landmarks correspond between subjects and time-points such that they have the same anatomical meaning and position on each bone, and the high density means that the femur model includes over 100 000 points. The density and correspondence of landmarks allows for the accurate analysis of differences in anatomical areas.

Figure 3: 3-dimensional change in the femur

This figure demonstrates the systematic change in femur shape with OA, represented as a position along a vector of OA shape scaled such that (-1) and (+1) represent the mean non-OA and mean OA shapes respectively.

Figure 4: Percentage change in bone area (tAB) for medial regions in OA and non-OA groups.

The graph shows percentage change from baseline for OA and non-OA groups, error bars are 95% confidence interval. All changes were highly significant (p<0.0001).
across the bony surfaces. Using this method, we have confirmed that height explains the majority of variance in bone area, confirming an allometric relationship between body and joint size.

Radiographic measures of OA, derived from a single radiographic projection, appear to account for only a small amount of variation in 3D knee OA total bone area\(^{[18]}\). Further studies have shown that changes in bone area discriminate people with OA from controls, and are more responsive than the current and impending standards for assessing OA progression\(^{[19]}\), see Figures 4 and 5.

In a further analysis of data in a subset of 352 participants from the Osteoarthritis Initiative, responsiveness of bone area change was compared with change in radiographic joint space width (JSW) and MRI cartilage thickness over a 2-year period. Responsiveness measured by the standardized response mean (SRM) at 12 months for bone area was 0.66, for JSW it was 0.19 and for cartilage thickness, 0.28. The increased sensitivity of the method is, in part, due to the improved repeatability. AAM segmentations were highly repeatable, with CoVs of less than 1%, compared with around 5% for JSW. Cartilage thickness CoV has been reported from the group performing the measures at around 2–3%. In a clinical trial expected to produce a 50% effect with a double-sided, 80% power, \(L=0.05\) design, then the cohort size assuming a 1-year trial would be as follows, JSW: 1298 patients per group, cartilage thickness: 459 patients, bone area: 149 patients, see Table 1 for a summary of these comparisons.

Table 1: Responsiveness and typical cohort size for various methods with a 50% effect size

<table>
<thead>
<tr>
<th>METHOD</th>
<th>TYPICAL COV</th>
<th>TYPICAL COHORT SIZE FOR 80% POWER, 50% EFFECT SIZE, (L=0.05) STUDY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radiographic JSW</td>
<td>5%</td>
<td>1298</td>
</tr>
<tr>
<td>MR cartilage thickness</td>
<td>2-3%</td>
<td>459</td>
</tr>
<tr>
<td>MR bone area</td>
<td>1%</td>
<td>149</td>
</tr>
</tbody>
</table>

Another recent large-scale study using MRI data from 600 subjects enrolled in the Osteoarthritis Initiative\(^{[20]}\) looked at longitudinal validation of bone area measurement and 3D shape as biomarkers for knee OA progression. This study showed that greater increases in bone area and shape markers over 24 months in knees with mild-to-moderate radiographic OA are associated with increased likelihood of clinically relevant progression consisting of a combination of radiographic and symptomatic progression over 48 months. The change in bone found in these studies provides an exciting new window on pathogenesis of the disease, and suggests that bone can now provide a new focus for clinical trials.

(A) shows the regions used in this study, displayed on the mean shape for each bone. MF, medial femur; LF, lateral femur; MT, medial tibia; LT, lateral tibia; MedPF, medial trochlear femur; LatPF, lateral trochlear femur; MP, medial patella; LP, lateral patella. The MF/MedPF and the LF/LatPF boundaries were defined as a line on the bone corresponding to the anterior edge of the medial or lateral meniscus in the mean model. The MedPF/LatPF boundary was defined as the centre of the trochlear groove in the mean model.

(B) shows schematic results for an active appearance model fit to 4 different femurs; each bone surface is fitted with a dense set of landmarks during auto-segmentation, which corrects for individual shape differences. It is impractical to display the actual density of the model, for example, the femur model includes over 100 000 points. Location of 4-year change for the non-OA group is displayed in (C) and for the OA group in (D). Areas which increase in size more than measurement error are coloured red; those with a similar decrease are coloured blue.
**Measurement** of Bone Marrow Lesions

MRI has highlighted the whole-organ nature of the OA process and in particular has demonstrated highly prevalent bone marrow lesions (BMLs). BMLs, defined as ill-delineated regions of hyper-intensity in fat suppressed MR images, are comprised of histological abnormalities such as fibrosis, necrosis, trabecular bone abnormalities and microfractures[21].

Factors which affect the formation of bone marrow lesions (BML) in knee osteoarthritis (OA) are poorly understood. They have been associated with both pain and progressive compartment-specific cartilage loss[22]. The study of BMLs is typically conducted using semi-quantitative methods such as WORMS or MOAKS, which do not provide good spatial information of either cartilage or BML lesions.

This is the method provided by Imorphics in the work by David Felson (for example[23]). Using accurate quantitative image analysis with confirmatory individual participant analysis, we have demonstrated for the first time the very strong 3D spatial relationship between OA BMLs and severe cartilage loss in the femur and tibia of knee OA (see Figure 6), both at individual and group level[24].

**Meniscus shape** and volume

The meniscus is crucial to the normal functioning of the knee, and damage or compromise to the meniscus is an important component in the development of knee osteoarthritis[25]. Quantitative measurement of the damage to the meniscus is likely to serve as a useful biomarker of OA progression[26].

In principle, the meniscus is a simple shape, however, damage to the meniscus may appear as damage to meniscal volume, extrusion of the meniscus, or a general failure of meniscal competence, resulting in the spreading of the surface.

Image analysis using statistical shape and appearance models enables accurate quantification of tissue morphology, and provides a consistent 3-dimensional framework. This framework can be utilized for highly reproducible measurement across a cohort of patients, and also to compare the spatial locations of different tissue morphologies.

We use manual segmentation of BMLs in the MR images, followed by consistent subdivision into anatomical regions using the 3D auto-segmented bone surface.

---

**Figure 6:** Spatial comparison of summary population values for denuded cartilage with typical BML volume

Top row: Denudation figures show the percentage of the population who have denuded cartilage at each point on the bone surface (scale in % shown in legend).

Bottom row: Average OA BML volumes.

**Figure 7:** Meniscal measurement strategies

(a) shows (from left to right) total meniscus volume, sections used for meniscal height (thickness) and extruded volume. (b) illustrates the method for measuring cartilage thickness from the corresponding points, using a sagittal slice through the bone and meniscus. From each correspondence point a measurement is taken normal to the surface. Meniscal thickness is shown as a green arrow.
We have employed statistical shape modelling to study a number of potential measures of meniscal deterioration within a one-year period (Figure 7). Additionally, we have used statistical models to visualize the areas of the menisci which undergo most change\cite{27}. Measures of meniscal volume and meniscal extrusion are very noisy, due to the many shapes which the damaged meniscus may adopt.

The most promising measure of meniscal change from our study is the meniscal window, measured either as an area or as a proportion of the cartilage plate. This measure is very responsive and should be relatively easy to perform for research groups, even without access to specialist 3D measurement.

Additionally, measurements of meniscal height over the tibia, used in a similar manner to our method for articular cartilage thickness measurement, appears to provide a promising measure of change.

**A note on MRI Imaging Sequences**

The imaging sequences we employ are used in standard clinical assessment of the bone and cartilage of the knee, and are also used in the SQ scoring methods such as WORMS, MOAKS and BLOKS:

- For cartilage, bone shape and meniscus measurements, the MR images should be high-resolution 3D images such as DESS (Siemens) or WatSC (Philips), or the GE equivalent.
- For bone marrow lesions we use T2 TSE or STIR images.

**Conclusions**

The use of MRI techniques to investigate tissue pathology has become increasingly widespread in osteoarthritis (OA) research. Semiquantitative assessment of the joints by expert interpreters of MRI data is a powerful tool that can increase our understanding of the natural history of this complex disease.

Several reliable and validated semiquantitative scoring systems (e.g. BLOKS, WORMS, MOAKS) now exist and have been applied to large-scale, multicentre, cross-sectional and longitudinal observational epidemiological studies. However, these approaches are time-consuming and require expert MSK radiologists to reduce inter-operator variation.

Compared to manual measurements and semiquantitive scoring, the use of automated quantitative measurements can increase the reproducibility of the measurements made, thereby increasing their sensitivity to change. In addition, these automated measurements are usually rapid to compute in comparison to manual methods, or else they may be run on batches of images unsupervised.

Perhaps, even more importantly, these automated methods can be used to produce imaging biomarkers. These novel measures such as the quantification of shape difference or the highly reproducible determination of regional area on complex 3D surfaces, may not be produced by human interaction and are beginning to yield fascinating insight into the disease process.