Disease severity classification using quantitative magnetic resonance imaging data of cartilage in femoroacetabular impingement

Henn LL
Disease severity classification using quantitative magnetic resonance imaging data of cartilage in femoroacetabular impingement

Lisa L. Henn\textsuperscript{a}, John Hughes\textsuperscript{b}\textsuperscript{*}, Eleena Iisakka\textsuperscript{a}, Jutta Ellermann\textsuperscript{c}, Shabnam Mortazavi\textsuperscript{c,d}, Connor Ziegler\textsuperscript{d}, Mikko J. Nissi\textsuperscript{c,d}, and Patrick Morgan\textsuperscript{d}

Femoroacetabular impingement (FAI) is a condition in which subtle deformities of the femoral head and acetabulum (hip socket) result in pathological abutment during hip motion. FAI is a common cause of hip pain, and can lead to acetabular cartilage damage and osteoarthritis (OA). For some patients with FAI, surgical intervention is indicated, and it can improve quality of life and potentially delay the onset of OA. For other patients, however, surgery is contraindicated because significant cartilage damage has already occurred. Unfortunately, current imaging modalities (x-rays and conventional MRI) are subjective and lack the sensitivity to distinguish these two groups reliably.

In this paper we describe the pairing of T2* mapping data (an investigational, objective MRI sequence) and a spatial proportional odds model for surgically obtained ordinal outcomes (Beck’s scale of cartilage damage). Each hip in the study is assigned its own spatial dependence parameter, and a Dirichlet process prior distribution permits clustering of said parameters. Using the fitted model, we produce a six-color, patient-specific predictive map of the entire acetabular cartilage. Such maps will facilitate patient education and clinical decision-making. Copyright © 2000 John Wiley & Sons, Ltd.

Keywords: Areal model; Arthroscopy; Bayesian composite likelihood; Copula; Dirichlet process; Hip; Osteoarthritis; Personalized medicine; Proportional odds; T2*

1. Introduction

Femoroacetabular impingement (FAI), a common cause of hip pain, is a condition characterized by abnormal peri-articular morphology that results in pathological abutment between the head–neck junction of the femur and the acetabular rim ([1, 2]). FAI has been shown to cause labral (edge, rim) and chondral (of or relating to cartilage) lesions and is a strong
risk factor for osteoarthritis ([3, 4, 5, 6]). For FAI patients with symptoms unresponsive to non-operative management, joint preservation surgery can be considered if cartilage damage is not significant ([7, 8, 9]). But joint preservation procedures are contraindicated for patients with moderate to advanced cartilage changes because more severe cartilage abnormality is associated with early conversion to total hip replacement ([10, 11]). Unfortunately, moderate cartilage damage can be challenging to diagnose ([12]). Radiographic evaluation with use of Tönnis grading (assignment of an ordinal score in {0, 1, 2, 3, 4} based on inspection) is the standard of care but has been shown to have poor inter-observer reliability ([13, 14, 15]). Magnetic resonance (MR) image evaluation seems a sensible alternative, but the accuracy of MR imaging and MR arthrography for detecting chondral damage in FAI is poor ([12, 16, 17, 18, 19, 20]).

The identification of cartilage damage in FAI can be difficult owing to the pattern of damage particular to the condition ([21]). In FAI, cartilage damage is frequently limited to the acetabulum and may initially occur deep within the tissue as a debonding of articular cartilage from acetabular bone ([3]). Since the superficial cartilage remains intact and traditional MR imaging is best suited for revealing damage at the articular surface, this pattern of damage hinders diagnosis. Thus investigators have turned to quantitative MR mapping techniques such as delayed gadolinium-enhanced MR imaging of cartilage ([22, 23, 24, 25]) and T2 mapping ([26, 27, 28, 29]). Delayed gadolinium-enhanced imaging is the most widely applied technique, but it can be time consuming and difficult to perform, and, due to limited resolution, currently does not allow segmentation of femoral and acetabular cartilage ([30]). Furthermore, use of gadolinium agents is contraindicated in patients with limited renal function. T2 relaxation time measurements ([26, 27, 28, 29]) and, more recently, T2* mapping have also been reported for the hip ([31, 32, 33, 34]). T2* has the advantages that it (1) can be acquired quickly, (2) does not require contrast material (which must be injected in the joint or intravenously), and (3) has sufficient resolution to differentiate between femoral and acetabular cartilage.

The aim of our study was to determine whether quantitative T2* mapping can be used for routine cartilage assessment in FAI. To do this, we first developed an anatomically precise technique for extracting T2* data. Then we compared extracted and aggregated T2* data with a surgical gold standard. To link T2* measurements to the cartilage’s surgically revealed condition, we applied a model that accounts for the spatial arrangement of the measurements in a given hip, and assigns to each hip a spatial dependence parameter, the magnitude of which governs the degree of dependence among the outcomes. We used a Dirichlet process prior distribution to permit the dependence parameters to take on only a few similar values. This permitted the data to select a spatial model that was neither too flexible nor too restrictive. Our model’s predictive performance was superior to that of a non-spatial model.

The remainder of the article is organized as follows. In Section 2 we describe the collection of our T2* data and corresponding arthroscopic data. In Section 3 we describe the statistical models and apply them to the data described in Section 2. In Section 4 we apply our fitted models to unaggregated patient T2* data, and produce a patient-specific map of (predicted) cartilage health. We conclude with a discussion in Section 5.

2. Collection of T2* Data and Corresponding Arthroscopic Data

2.1. Patients

We collected our data between February, 2010 and March, 2012. The study group included 26 patients (28 hips) who exhibited clinical and radiographic signs of FAI, showed no evidence of osteoarthritis, underwent the study imaging protocol, were diagnosed with a labral tear, and subsequently underwent hip arthroscopy after conservative treatment failed. The clinical diagnosis of FAI was established by the presence of moderate to severe and persistent hip or groin pain that limited activity and worsened with flexion activity, and positive impingement sign (i.e., sudden pain at 90° hip flexion with adduction and internal rotation, or with extension and external rotation). Radiographic confirmation of FAI was based on findings such as α angle (angle between femoral neck and line through center of the head and point where the head becomes aspherical) greater than 50°, pistol grip deformity (non-spherical femoral head), coxa profunda (a too-deep
socket), and acetabular retroversion (the mouth of the acetabulum inclines posterolaterally). Exclusion criteria included osteoarthritis as evidenced by Tönnis grade $> 1$ ([15]), previous hip surgery, or diagnosis of other abnormalities to which the patient’s hip pain could be attributed. All patients were examined by Dr. Morgan, who specializes in hip arthroscopy, and evaluated with standardized radiographs per published protocol ([35]).

There were nineteen female and seven male patients. There were twenty female hips and eight male hips. The mean age for all patients was 29.0 (range 12–53). The mean age for females was 28.9 (range 12–46). The mean age for males was 29.1 (range 16–53). All hips had a Tönnis grade of 0 or 1, indicating at most mild signs of osteoarthritis.

2.2. Imaging

A 3T clinical imaging protocol (Trio; Siemens Medical Solutions, Erlangen, Germany) was used. The protocol required approximately 45 minutes to complete, with the T2* data obtained during the final seven minutes to control for time dependence of T2* values after unloading ([34]). T2* maps were generated using software provided by the scanner vendor (Mapit; Siemens Medical Solutions).

Region of interest (ROI) analyses were performed by Dr. Ziegler, a second-year orthopedic resident, who was blinded to the patients’ clinical information. Acetabular orientation was standardized on sagittal images by using a line passing through the center of the femoral head, perpendicular to the transverse acetabular ligament, defining the 12-o’clock position (Figure 1). Case regions of interest were defined in the anterosuperior acetabulum because this area has the highest reported incidence of damage in patients with FAI ([3, 16, 36]). Using the image processing application OsiriX ([37]), acetabular cartilage in this region was divided into five ROIs between the 12-o’clock position and the chondrolabral junction (Figure 1). This was done for three consecutive sagittal sections, yielding a total of fifteen case ROIs. Four control ROIs were defined in the posteromedial acetabulum, where articular cartilage damage is infrequent in FAI. Note that this landmark-based extraction resulted in ROIs comparable between the patients even though the volume (and number of voxels) varied from patient to patient.

2.3. Arthroscopy

To evaluate the utility of T2* for assessing cartilage damage, we needed a reference assessment. For our study, the reference assessment was obtained through arthroscopic surgery. All arthroscopic examinations were performed by Dr. Morgan, who was blinded to the T2* data. Dr. Morgan was presented with a patient-specific, flattened anatomical map of the acetabulum (an example of which is shown in Figure 2), on which simple, obvious bony landmarks could be co-located during surgery. Once located, individual ROIs were measured relative to a flexible probe measuring 2 mm in diameter, which served as a ruler. Dr. Morgan recorded his surgical findings on the patient-specific acetabular projection. A modified Beck scale (described in Table 1, [3]) was used to characterize the degree of articular cartilage damage. Both case and control ROIs were assessed (532 ROIs in all, 19 per hip).

The data are summarized in Table 2. Note that the T2* values for a given ROI were aggregated by taking the sample mean over all voxels in the ROI, and so the sample quantities given in the table are for samples of sample means. Box plots of T2* by Beck score are shown in Figure 3.
3. Development and Validation of the Predictive Models

We assume that Beck scores are ordinal multinomial outcomes that satisfy the proportional odds assumption ([38]). Specifically, we suppose that

\[
\gamma_k(x_{ij}) = \Pr(Y_{ij} \leq k \mid x_{ij}) = \frac{\exp(\alpha_k - \beta' x_{ij})}{1 + \exp(\alpha_k - \beta' x_{ij})}
\]

\[(i = 1, \ldots, 28) \ (j = 1, \ldots, 19) \ (k = 1, \ldots, 5),\]

where \(Y_{ij}\) is the Beck score for the \(j\)th ROI in the \(i\)th hip, \(k\) is the Beck category, \(\alpha_k\) is the intercept (or threshold) for category \(k\), non-decreasing from its predecessor, \(x_{ij}\) is a vector of predictors associated with ROI \(ij\), and \(\beta\) are regression coefficients. Equivalently,

\[
\text{logit} \ \gamma_k(x_{ij}) = \alpha_k - \beta' x_{ij}
\]

or

\[
\pi_k(x_{ij}) = \Pr(Y_{ij} = k \mid x_{ij}) = \begin{cases} 
\gamma_1(x_{ij}) & k = 1 \\
\gamma_k(x_{ij}) - \gamma_{k-1}(x_{ij}) & k \in \{2, 3, 4, 5\} \\
1 - \gamma_5(x_{ij}) & k = 6.
\end{cases}
\]

Having no subscript on \(\beta\) implies that the predictors have the same effects for all categories. This assumption and the use of the logit link function together characterize the proportional odds model, the most common of the so-called cumulative link models for ordinal outcomes.

We also assume that Beck scores are dependent within a patient. To reflect the physical conditions at each ROI, we formulate an explicit model for the similarity among spatially proximate observations. Our spatial model incorporates a Dirichlet process prior distribution to describe similarities in the spatial dependence parameters from hip to hip. This results in shrinkage of the estimated parameter values toward common cluster means, which represents a data-driven approach to selecting a spatial model that is neither too restrictive nor too flexible. We also consider, for the sake of comparison, a model that accommodates within-patient dependence in a crude, non-spatial fashion.

3.1. Spatial Modeling of Beck Scores

We might well expect spatially proximate Beck scores to be similar, and spatially distant scores to be (at least nearly) independent. In other words, we might expect the scores to exhibit positive spatial dependence, which we can accommodate by applying a spatial model. More specifically, since each Beck score was assigned to all voxels in the corresponding ROI, we elected to apply an areal model, i.e., a model for spatially aggregated data. In areal modeling, the spatial proximity among observations is expressed by way of the adjacency structure among the areal units, in this case the ROIs, from which the observations were obtained. The adjacency structure is used to define the spatial dependence model for the outcomes. In this analysis, we count as adjacent to an ROI those ROIs immediately anterior and posterior to the ROI as well as those ROIs in the same position in adjacent sagittal sections, for a total of four neighbors for observations in the interior of the measurement region. A trace on Figure 1 of a neighborhood so defined would take the shape of a plus sign.

Our areal model is an extension of the copCAR model, which was developed by Hughes ([39]). The copCAR model
employs a Gaussian copula exhibiting the well-known proper conditionally autoregressive (CAR) dependence structure ([40]) for areal data. In the next section we briefly describe the copCAR model; a more extensive treatment appears in Hughes ([39]).

3.1.1. The copCAR Model The copCAR model is a Gaussian copula regression model (GCRM) that employs a proper CAR dependence structure. Specifically, the (unscaled) precision matrix for the proper CAR model is

$Q = D - \rho A,$

where $D$ is diagonal with $D_{ii}$ equal to the number of areal units that are adjacent to the $i$th areal unit, $\rho \in [0, 1)$ is the spatial dependence parameter, and the binary matrix $A$ embodies the adjacency structure among the areal units: $A_{ij} = 1$ if areal units $i$ and $j$ are adjacent, $A_{ij} = 0$ otherwise. The parameter $\rho$ is best thought of as a range parameter, with values closer to 1 leading to longer-range, and thus stronger, spatial dependence.

This precision matrix is used to construct the CAR copula,

$$C_{Q^{-1}}(u) = \Phi_{Q^{-1}}\left\{\Phi_{\sigma_1^2}^{-1}(u_1), \ldots, \Phi_{\sigma_n^2}^{-1}(u_n)\right\},$$

which has copula density

$$c_{Q^{-1}}(u) \propto |Q|^{1/2} |\Sigma|^{1/2} \exp \left\{-\frac{1}{2} z'(Q - \Sigma^{-1}) z\right\},$$

where the $\sigma^2$ are the diagonal elements of $Q^{-1}$, $\Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2)$, $\Phi_{\sigma_i^2}^{-1}$ denotes the quantile function for a normal random variable with mean zero and variance $\sigma_i^2$, and $z_i = \Phi^{-1}\{F_i(y_i | \beta)\}$ with $F_i$ being the cdf for the $i$th outcome. Note that the CAR variances are identifiable since they are entirely determined by $\rho$ and the adjacency structure. The variances can be approximated efficiently as described by Hughes ([39]), and retaining them (as opposed to inverting $Q$ and converting it to a correlation matrix) permits faster computation for larger datasets.

The spatial association among the observations is thus described using the precision matrix $Q$, which incorporates the adjacency structure among the areal units. The marginal distributions of the observations can be described flexibly through the distribution functions $F_i$, and they are “linked” to the CAR copula using the probability integral transform.

3.1.2. Bayesian Composite Likelihood Inference for copCAR When the marginal distributions are discrete, the joint distribution function of the copula, such as shown in (3), is uniquely defined only on the support of the marginals, and the dependence between a pair of random variables depends on the marginal distributions as well as on the copula. This was described by Genest and Nešlehová ([41]), who explained the implications and warned that, for discrete data, “modeling and interpreting dependence through copulas is subject to caution.” However, Genest and Nešlehová ([41]) went on to say that one can still interpret copula parameters as dependence parameters and that estimation of copula parameters is often possible using fully parametric likelihood-based methods.

When the outcomes are discrete, the likelihood is

$$L(\theta | y) = \sum_{j_1=0}^1 \cdots \sum_{j_n=0}^1 (-1)^k C_R(u_{1j_1}, \ldots, u_{nj_n}),$$

where $k = \sum_{i=0}^n j_i$, $u_{i0} = F_i(y_i)$, and $u_{i1} = F_i(y_i^-) \equiv \lim_{y_i \to y_i^-} F_i(y)$. Computing (5) is generally infeasible because the number of summands grows exponentially. Moreover, the multivariate normal distribution function can be numerically unstable in high dimensions. These challenges necessitate approximation. We use a composite likelihood in place of the true likelihood.

In his seminal paper, Lindsay ([42]) described a composite likelihood as a “likelihood type object formed by adding together individual component log likelihoods.” The component log-likelihoods are formed for subsets of the response vector. To analyze our FAI data we used a composite marginal likelihood ([43]), where the objective function was a
product of pairwise likelihoods:

\[
L_{\text{CL}}(\theta \mid y) = \prod_{i,j \in \{1, \ldots, n\}, i \neq j} \frac{1}{2} \sum_{j_1=0}^{1} \sum_{j_2=0}^{1} (-1)^k C_{ij}(u_{ij_1}, u_{jj_2}),
\]

where \( k = j_1 + j_2 \) and \( C_{ij} \) denotes the bivariate Gaussian copula with correlation matrix

\[
R^{ij} = \begin{pmatrix}
R_{ij} & R_{ii} \\
R_{ji} & 1
\end{pmatrix},
\]

with \( R \) being the correlation matrix for the corresponding \( n \)-copula. This implies the log composite likelihood

\[
\ell_{\text{CL}}(\theta \mid y) = \sum_{i \in \{1, \ldots, n-1\}} \log \left\{ \frac{1}{2} \sum_{j_1=0}^{1} \sum_{j_2=0}^{1} (-1)^k \Phi_{R^{ij}}(z_{ij_1}, z_{jj_2}) \right\},
\]

(6)

where \( z_{*0} = \Phi^{-1}\{F_{\theta}(y_*)\} \) and \( z_{*1} = \Phi^{-1}\{F_{\theta}(y_* - 1)\} \).

For the CL approach to be valid in a Bayesian setting, the resulting posterior distribution must be integrable. Since we use proper prior distributions, it is sufficient, following the argument set forth by Ribatet et al. ([44]), that for each term in the composite likelihood there exists a finite \( b_\theta \) such that \( \sup_\theta f(y \in A_\theta \mid \theta) \leq b_\theta \). Then

\[
\int L_{\text{CL}}(\theta \mid y) p(\theta) \, d\theta = \int \prod_{\epsilon \in \mathcal{I}} f(y \in A_\epsilon \mid \theta)^{b_\theta} p(\theta) \, d\theta \\
\leq \prod_{\epsilon \in \mathcal{I}} b_\theta^{b_\theta} < \infty,
\]

where \( p(\theta) \) denotes a prior distribution. Inspection of (6) reveals that \( b_\theta = 1 \) is a suitable bound. The posterior distribution is thus well defined and can be used in the Bayesian analyses described below. Ribatet et al. ([44]) also addressed the question of convergence when a composite likelihood is used in MCMC for Bayesian inference.

The pairwise composite likelihood is mis-specified since it assumes that any two pairs of observations are independent. To make this assumption is to consider the data to be more informative than they actually are, and optimistic inference is the likely result. To mitigate this possibility, we apply a curvature correction. The correction, described by Chandler and Bate ([45]), approximates the quadratic component of the exact log-likelihood by evaluating the mis-specified objective function at a proxy value for the parameter vector: \( \ell_{\text{curv}}(\theta \mid y) = \ell_{\text{CL}}(\theta_{\text{adj}} \mid y) \), where \( \theta_{\text{adj}} = \hat{\theta}_{\text{CL}} + B(\theta - \hat{\theta}_{\text{CL}}) \), with \( B \) being a matrix to be described shortly and \( \hat{\theta}_{\text{CL}} \) being the estimate resulting from optimization of \( \ell_{\text{CL}} \).

The matrix \( B \) ensures the expected Hessian of the modified parameter vector equals the Godambe information of the original vector. That is to say,

\[
B = M^{-1} M_A,
\]

where

\[
M_A M_A' = -H(\theta_0) J(\theta_0)^{-1} H(\theta_0)
\]

and

\[
M' M = H(\theta_0).
\]

In the preceding, \( H(\theta_0) \) is the Hessian of the original parameter vector for \( \ell_{\text{CL}}(\theta_0 \mid y) \) and \( J(\theta_0) \) is the variance of the score function for the original parameter vector. The technique thus effectively modifies the information of the original parameter vector to yield the robust variance estimate needed in the presence of mis-specification.
In practice, \( \theta_0 \) is unknown. An estimate of the matrix \( \mathcal{H}(\theta_0) \) is \( \mathcal{H}(\hat{\theta}_{cl}) \), which is easy to obtain. We estimate \( \mathcal{F}(\theta_0) \) using a parallel parametric bootstrap (\([39]\)). As Chandler and Bate (\([45]\)) noted, the matrix square roots \( \mathcal{M}_A \) and \( \mathcal{M} \) are not unique. To preserve any asymmetry that may be present, we use a singular value decomposition to compute \( \mathcal{M}_A \) and \( \mathcal{M} \).

Ribatet et al. (\([44]\)) also noted that an MCMC algorithm based on the curvature-corrected composite likelihood must be shown to converge to the correct target distribution. They argued that the conditions for Theorem 7.2 of Robert and Casella (\([46]\)), which establish detailed balance for the Metropolis–Hastings algorithm, are satisfied for the composite likelihood with curvature correction. The proposal distribution meets the criteria given in (7.3) of Robert and Casella (\([46]\)). Theorem 7.2 of Robert and Casella (\([46]\)) builds upon Theorem 6.46 of that work and requires the target to be a distribution. As established above, although the composite likelihood is not the true likelihood, it nevertheless is integrable, and so the posterior distribution is well defined.

We analyzed our FAI data in a Bayesian setting using the Metropolis–Hastings algorithm. We used the same predictors as in the most parsimonious model from Section 3.2, in order to compare the two approaches. In our analysis, each hip has a potentially unique association parameter. We re-parameterized the thresholds \( \alpha_k \) for \( k > 1 \) to the form \( \Delta_k = \alpha_k - \alpha_{k-1} \) to ensure, in conjunction with the prior and proposal distributions, that the corresponding \( \alpha_k \) parameter proposals are non-decreasing. We applied independent \( \mathcal{N}(0, 10^6) \) prior distributions to \( \alpha_1 \) and the elements of \( \beta \), and we applied independent Gamma(2, 1) prior distributions to the \( \Delta_k \) terms. The parameters of these gamma prior distributions place most of the mass in the range (0, 4), corresponding to results from earlier analyses. To propose \( \alpha_1 \) and \( \beta \) values, we used Gaussian random walks. For each of the \( \Delta_k \) parameters, we first proposed a value from a normal distribution centered at the log of the most recently accepted value. Then we exponentiated the value to at our proposed value for \( \Delta_k \). For all random walk proposals, we used the estimated asymptotic covariance sub-matrices from a fixed-effects proportional odds model fit, scaled to obtain acceptance rates in the range of 20%–50%. We ran the chain for 120,000 iterations, which resulted in small Monte Carlo standard errors (\([47]\)).

Recall that the association parameters \( \rho_i \) take values in \([0, 1]\). The curvature correction may yield adjusted values that fall outside this range. We obviated this problem by generating values on a transformed scale (described in 3.1.3).

Chandler and Bate (\([45]\)) note that the curvature correction “lacks invariance to reparameterization.” They advise avoiding excessively skewed log-likelihoods. Our pilot studies with a test case suggest that our chosen transformation performs satisfactorily in this respect.

### 3.1.3. Clustering of Spatial Association Parameters

In a preliminary spatial analysis of these data, Iisakka (\([48]\)) noted that a model with a single dependence parameter for all hips yielded implausible parameter estimates, while allowing a different dependence parameter for each hip led to an objective function that could not be optimized. We found that a Dirichlet process prior distribution provides a compromise between these two extremes, allowing \( \rho_i \) values to cluster and “borrowing strength” across hips.

We envision the dependence parameters, one for each hip, as draws from several members of a family of distributions. A Dirichlet process is a prior distribution for the mixing distribution of these family members. We therefore can think of a Dirichlet process as a “distribution of distributions,” and using it eliminates the need to know \emph{a priori} how many members of the family to use.

One characterization for the Dirichlet process is as a generalization of a beta distribution. A beta distribution can model proportions, the distribution of propensities to be in one of two categories. The Dirichlet distribution generalizes this idea to a mixture among \( t \) categories. The Dirichlet process generalizes further. It results when \( t \to \infty \). Each realization of the Dirichlet process is a Dirichlet distribution, discrete with probability one (\([49]\)). As Escobar and West (\([50]\)) point out, there is a positive probability of coincident values in each realization of the Dirichlet process. These characteristics enable the Dirichlet process to model an arbitrary number of clusters. Each realization contains a finite number of elements, describing (non-zero) proportions in each of a number of categories not specified ahead of time. Sufficient sampling
from a given realization inevitably results in repeated values. For us, this allowed multiple hips to have the same value of the spatial dependence parameter \( \rho \). The number of categories with non-zero proportions can change from realization to realization. Each category can be a parameter value. Alternately, each category can be an identifier for a member in a family of distributions, with the implication that each realization of the Dirichlet process describes a mixture of these distributions.

The Dirichlet process is specified by two attributes: a base distribution and a concentration parameter. As Neal (\cite{51}) writes, the Dirichlet process prior distribution can be specified as

\[
\begin{align*}
y_i | \theta_i & \sim F(\theta_i) \\
\theta_i | G & \sim G \\
G & \sim \mathcal{DP}(G_0, a),
\end{align*}
\]

for distribution \( F(\theta_i) \), base distribution \( G_0 \), and concentration parameter \( a \). The base distribution represents the prior expectation of the realization \( G \). The concentration parameter describes how closely the mass of a given realization concentrates around this expectation. For a measurable set \( A \), as \( a \to \infty \), \( G(A) \to G_0(A) \) (\cite{52}). The smaller the concentration parameter, the fewer clusters there are likely to be (\cite{53}).

Since we performed our MCMC analysis using the transformation \( \nu_{\rho_i} = \Phi^{-1}(\rho_i) \), our choice of base distribution was unimportant; we used the \( \mathcal{N}(0.85, 0.0225) \) distribution as the base distribution. The concentration parameter’s prior distribution was a Gamma(4, 1), which applies most of the mass in the range \((0, 10)\). Preliminary studies in which we fixed the concentration parameter to single-digit integer values showed, through the acceptance rates, that the data favor low values of \( a \). See Neal (\cite{51}) and Escobar and West (\cite{50}) for details on the Markov chain sampling methods for the Dirichlet process. With independent prior distributions, the Markov chains for certain association parameters struggled to propose acceptable candidate values. With the Dirichlet Process prior distribution, all estimated parameters showed excellent acceptance rates.

3.2. A Non-Spatial Model for Beck Scores

As described in Iisakka (\cite{48}), our non-spatial approach to modeling the intra-patient dependence of Beck scores used a random intercept. That is, we assumed the outcomes for a given hip to be independent conditional on a shared random effect:

\[
\text{logit}\; \gamma_k(x_{ij}, W_i) = \alpha_k - \beta' x_{ij} + W_i,
\]

where the \( W_i \) follow a spherical Gaussian distribution located at 0. In addition to T2*, we included age, weight, and sex as predictors, as well as quadratic and cubic terms for the continuous predictors and two-way interactions among all predictors. Both weight and T2* were centered at their mean values. Corresponding quadratic and cubic terms were calculated from these centered values. We fit each candidate model using the function \texttt{clmm()} of the R (\cite{54}) package \texttt{ordinal} (\cite{55}). Since the random intercepts imply an analytically intractable likelihood (the likelihood contains an integral over \( \mathbb{R}^{28} \)), the likelihood must be approximated. Function \texttt{clmm()} approximates the integral using the Laplace approximation (\cite{56}) or adaptive Gaussian–Hermite quadrature (\cite{57}). The number of quadrature points can be specified using the argument \texttt{nAGQ}. We found that ten quadrature points yielded sufficiently accurate maximum likelihood estimates. Two other arguments of interest are \texttt{link} and \texttt{threshold}. We initially chose the logit link function and flexible thresholds (i.e., the only constraint on the thresholds was that they be strictly increasing). The most parsimonious model among the random intercept models considered was determined by backward elimination.
3.3. Results

The coefficient and threshold estimates for both models are given in Table 3. The random intercept model is quadratic in both centered T2* and centered weight, has a significant effect for sex, and has a different effect of centered T2* for males and females. The inclusion of weight is justified because body mass index is known to be an independent risk factor for FAI ([58, 59]). More specifically, a decrease (increase) in T2* (weight) is associated with larger Beck scores, and males have larger Beck scores than females, on average. Males are considerably more likely to have cam-type FAI ([60]), which is more damaging, and young, active males with FAI tend to have more severe damage ([61, 62]). The random effects are necessary, for the independence model lacks fit (LRT; p value < 0.001). The estimated standard deviation of the random effects was 1.640, a significant fraction of the effective range for parameter estimates on the logit scale. This implies very strong within-hip dependence. We found no evidence against the proportional odds assumption, nor did we find that a different link function or structure for the thresholds (symmetric or equidistant) would be more appropriate.

[Table 3 about here.]

Figure 4 shows the estimated cumulative logits (as functions of centered T2* but without estimates of random effects) for females and males, with patient weights fixed at the sex-specific sample means (151.68 lbs and 180.75 lbs, respectively). These plots reveal the dramatic effect of sex.

[Figure 4 about here.]

For the areal model, we constructed the $\alpha_2$ to $\alpha_5$ values from $\alpha_1$ and $\Delta_k$ results at each iteration of the MCMC analysis, taking the means of the resulting vectors of values. In addition to the CL, we attempted analysis of the areal model using the continuous extension (CE) proposed by Denuit and Lambert ([63]) as well as the distributional transform (DT) described by Rüschendorf ([64]). In both cases parameter estimates were poor. In the CL fit, only centered T2* and the square of centered weight have intervals that exclude zero, making this spatial model more parsimonious than the random intercept model. The magnitudes of the effect estimates are smaller than for the random intercept model, most pronounced for centered T2*. Neighborhoods defined for this model group observations in the $3 \times 5$ case zone and in the $2 \times 2$ control zone. No neighbor relations bridge these two zones. This neighborhood structure adjusts the model similarly to an indicator of case or control observation. Indeed, when we added such an indicator variable to the random intercept model, the resulting parameter estimates hewed very closely to those of the areal model, as shown in Table 4. This finding, and the areal model’s relative parsimony, demonstrate the merit of adjusting for the spatial relationship among observations.

[Table 4 about here.]

The posterior mean for $a$ is approximately 1, indicating that the data strongly favor realizations featuring few distinct values of $\rho_i$. Similarly, the mean of the number of unique $\rho_i$ values in each iteration was approximately 1.5. All estimated $\rho_i$ values were at least as large as 0.965. A review of Beck scores on a hip-by-hip basis showed very strong association among scores within a given hip. In some cases, all measurement locations showed the same Beck score. Large $\rho_i$ are clearly appropriate for these data.

During a followup study we collected data for an additional 27 hips. We then applied both predictive models to the new data, and found that both models performed well. The results are shown in Table 5.

[Table 5 about here.]

4. Application to Patient T2* Data

In this section we will present and interpret two maps of predicted Beck scores for one patient from the study.
Figure 5 shows the patient’s predicted Beck map using the fits of the random intercept model and the spatial association model. We chose this color scheme for two reasons: (1) it is colorblind-friendly, and (2) the colors for adjacent Beck values are sufficiently different that one can easily distinguish the values, whereas a gradient would make it difficult to do so. We see from the map that this patient’s cartilage damage is significant in several regions, with large bands of red and blue present. There are, however, two large regions of relatively healthy tissue, colored in black and white.

Both models use centered predictors. Because the centered $T2^*$ effect is larger in absolute terms in the random intercept model than in the spatial model, a given departure from the average $T2^*$ value will tend to yield more extreme predictions in the random intercept model than in the spatial model. Thus, the spatial model predicts somewhat less damage (more moderate Beck scores) than the random intercept model in regions of negative centered $T2^*$ values. Where the centered $T2^*$ values are positive, the spatial model predicts somewhat greater damage (closer to 3 or 4) than the random intercept model. Both models broadly indicate the same regions as damaged, however, providing confirmation of the suitability of both models.

5. Discussion

Clinical MR imaging has significant limitations for providing the orthopedic surgeon with the most relevant information on cartilage integrity. The limitations are threefold. First is the inadequate visualization of a thin, approximately spherical 3D structure by an imaging format comprising consecutive two-dimensional slices. Even when data are acquired with three-dimensional sampling, the resulting images occupy a 3D Cartesian grid composed of imaging planes. The obliquity of planes for viewing can be variable, but the visualization of the cartilage is possible only when such planes intersect with the approximately spherical shape. Second, the grayscale intensities in clinical images do not provide information for quantitative cartilage assessment. Finally, such clinical images do not predict probability of disease for therapy stratification.

In this work we developed a predictive model incorporating the spatial relationship between observations and compared that model to a simple, non-spatial model. In the absence of a predictor for case or control zone, the only variables in the random intercept model that change over the 19 within-hip observations are centered $T2^*$ and its square. All other variables in the random intercept model are constant over a given hip. The results shown in Table 3 represent the best fit that can be made with only functions of $T2^*$ to capture the variation in Beck scores across a given hip. This fit overstates the effect of centered $T2^*$ and, by extension, the effect of its square.

In this particular analysis, we were able to adjust the random intercept model for case or control. Ordinarily, this information is unavailable. On the other hand, the neighborhood structure of the spatial model incorporates a de facto adjustment for case/control zone, or more generally, the similarity of observations to their neighbors. With a proper adjustment for this spatial dependence, we can better estimate the effect of centered $T2^*$ despite the fact that our sample does not capture the full continuum of change in $T2^*$. Thus, the spatial model estimates parameters of the model better than a predictive model that does not incorporate a spatial element in that the spatial model more closely reflects the true effect size of centered $T2^*$.

Given a patient’s sex, weight, and $T2^*$ data, our model can be used to produce predictive maps of the acetabular cartilage. The maps provide a non-invasive means of assessing the pattern, degree, and extent of articular cartilage damage, which can help clinicians to decide among non-operative management, joint preservation surgery, or joint replacement in FAI.

The Dirichlet process proved a suitable prior distribution for inference on the association parameters in this data set. Few continuous distributions can be parameterized intuitively such that the majority of the mass is clustered just to the left of 1. Some hips had log-likelihood values that dropped off steeply as the association parameter value decreased, whereas
for others the decline was nearly flat. The implication of this shape difference is that it would be difficult to find a single proposal distribution that would yield good MCMC acceptance rates for all hips.

One potential way to address this difficulty would be to define different proposal distributions for different groups of hips. But the small data set provides little information to suggest natural clustering. The Dirichlet process, by contrast, allows clustering to be informed by the data. Moreover, the mechanism by which values are sampled from realizations of the Dirichlet process facilitates good MCMC acceptance rates across all hips despite such a difference, providing more flexibility in choice of base distribution.

We also can develop analogous models for dichotomized outcomes, i.e., diseased or not. Such models necessarily discard information, namely, the degree of articular cartilage damage. For that reason, we recommend the full models presented here.

Two patients have data from both hips represented in this data set. One might consider that these two pairs of hips should receive the same adjustment for each member of the pair—the same random intercept or the same spatial association parameter, as the case may be. Incorporating this assumption had a negligible effect on the results.

Our model has the potential to inform surgeons of the extent of cartilage damage without resorting to invasive surgery. An important next step in the evaluation of this model’s classification ability would be to apply it to T2* data acquired from both damaged and undamaged hips. For cases in which surgery is performed, the assessment of the need of the procedure post-surgery by a surgeon blinded to the model’s prediction would shed light on how well the model might help to avoid unnecessary procedures. The work presented here represents a first, promising step for a viable new approach to FAI evaluation.

Acknowledgement

The first and second authors are indebted to Jim Hodges for many helpful discussions.

References


Figure 1. Selection of ROIs in acetabulum. (A) Gradient recalled echo image of hip, sagittal view. (B) Transverse ligament (T) used to find 12-o’clock position (12); angular guides for case ROIs in anterior superior labrum between 12-o’clock position and chondrolabral junction (CLJ); and angular guides for control ROIs (clockwise from 12-o’clock). (C) Magnified portion of panel B shows the dark line that delineates the boundary between acetabular and femoral cartilage.
Figure 2. A flattened imaging slice and ROI map of acetabular cartilage.
Figure 3. Box plots of aggregated T2* values by Beck score.
Figure 4. Estimated cumulative logits for females and males, as functions of centered $T^*_2$ and with patient weights fixed at the sex-specific means. Points on the plots correspond to $\gamma_k$ values from (1) for a given value of centered $T^*_2$. For example, for females at the mean $T^*_2$ value (centered $T^*_2 = 0$), $\gamma_1$ through $\gamma_5$ values are 0.387, 0.878, 0.984, 0.999, and 1.000 respectively.
Figure 5. Predicted Beck maps of the patient’s acetabulum for (a) the random-intercept model and (b) the spatial model.
**Table 1.** The modified Beck scale.

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>normal</td>
<td>macroscopically sound cartilage</td>
</tr>
<tr>
<td>2</td>
<td>early changes</td>
<td>softening; fibrillation; cartilage remains adhered to bone</td>
</tr>
<tr>
<td>3</td>
<td>debonding</td>
<td>loss of fixation to bone; “carpet” phenomenon</td>
</tr>
<tr>
<td>4</td>
<td>cleavage</td>
<td>loss of fixation to bone; frayed edges; thinning of cartilage; flap</td>
</tr>
<tr>
<td>5</td>
<td>defect, fibrous base</td>
<td>full-thickness loss of cartilage; thin fibrous tissue-covered base</td>
</tr>
<tr>
<td>6</td>
<td>defect, eburnated base</td>
<td>full-thickness loss of cartilage; base of eburnated bone</td>
</tr>
</tbody>
</table>
Table 2. Summaries, by Beck score, of the aggregated T2* values from our data set.

<table>
<thead>
<tr>
<th>Beck Score</th>
<th>Frequency</th>
<th>Sample First Quartile</th>
<th>Sample Median</th>
<th>Sample Third Quartile</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>172</td>
<td>22.1</td>
<td>28.7</td>
<td>33.5</td>
</tr>
<tr>
<td>2</td>
<td>160</td>
<td>16.7</td>
<td>21.1</td>
<td>27.6</td>
</tr>
<tr>
<td>3</td>
<td>112</td>
<td>16.1</td>
<td>20.0</td>
<td>23.9</td>
</tr>
<tr>
<td>4</td>
<td>59</td>
<td>14.8</td>
<td>18.9</td>
<td>22.7</td>
</tr>
<tr>
<td>5</td>
<td>21</td>
<td>13.8</td>
<td>17.3</td>
<td>19.5</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>14.2</td>
<td>16.4</td>
<td>18.5</td>
</tr>
<tr>
<td>Covariate</td>
<td>Random intercept model</td>
<td>Areal model with Dirichlet Process prior distribution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>------------------------</td>
<td>-------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Estimate</td>
<td>95% CI</td>
<td>Posterior Mean</td>
<td>95% HPD Interval</td>
</tr>
<tr>
<td>T2*</td>
<td>-0.146 (-0.1806, -0.1118)</td>
<td>-0.0534 (-0.0819, -0.0249)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(T2*)²</td>
<td>-0.005 (-0.0078, -0.0018)</td>
<td>-0.0014 (-0.0033, 5e-04)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weight</td>
<td>0.015 (0.0076, 0.0369)</td>
<td>0.0113 (-0.0037, 0.0255)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weight²</td>
<td>0.001 (0.0003, 0.0015)</td>
<td>6.522E-4 (2.241E-4, 1.095E-3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex (Male)</td>
<td>1.534 (-0.0026, 3.070)</td>
<td>0.7403 (-0.2668, 1.759)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2* × Sex</td>
<td>-0.089 (-0.1477, -0.0301)</td>
<td>-0.0366 (-0.0841, 0.0086)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>α₁</td>
<td>-0.517 (-1.538, 0.503)</td>
<td>-0.1145 (-0.7865, 0.5347)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>α₂</td>
<td>1.912 (0.8774, 2.947)</td>
<td>1.4357 (0.7489, 2.1273)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>α₃</td>
<td>4.037 (2.953, 5.122)</td>
<td>2.8555 (1.974, 3.7099)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>α₄</td>
<td>6.626 (5.349, 7.903)</td>
<td>4.6793 (3.4184, 6.0124)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>α₅</td>
<td>8.636 (7.130, 10.141)</td>
<td>6.0494 (4.4793, 7.6323)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4. Coefficient and threshold estimates for random intercept model with adjustment for observation location. Areal model results presented again for comparison.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Adj. Random Intercept Model</th>
<th>Spatial Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>95% CI</td>
</tr>
<tr>
<td>T2*</td>
<td>-0.0426</td>
<td>(-0.0824, -0.0027)</td>
</tr>
<tr>
<td>(T2*)^2</td>
<td>-0.0017</td>
<td>(-0.0013, -0.0048)</td>
</tr>
<tr>
<td>Weight</td>
<td>0.0210</td>
<td>(-0.0067, 0.0488)</td>
</tr>
<tr>
<td>Weight^2</td>
<td>0.001</td>
<td>(0.0004, 0.0019)</td>
</tr>
<tr>
<td>Sex (Male)</td>
<td>1.644</td>
<td>(-0.2628, 3.552)</td>
</tr>
<tr>
<td>T2* × Sex</td>
<td>-0.1093</td>
<td>(-0.1717, -0.0469)</td>
</tr>
<tr>
<td>Control ROI</td>
<td>-4.9240</td>
<td>(-5.884, -3.964)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Threshold</th>
<th>Estimate</th>
<th>95% CI</th>
<th>Mean</th>
<th>95% HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>α1</td>
<td>-1.097</td>
<td>(-2.364, 0.171)</td>
<td>-0.1145</td>
<td>(-0.7865, 0.5347)</td>
</tr>
<tr>
<td>α2</td>
<td>1.903</td>
<td>(0.6219, 3.1844)</td>
<td>1.4357</td>
<td>(0.7489, 2.1273)</td>
</tr>
<tr>
<td>α3</td>
<td>4.374</td>
<td>(3.0396, 5.7085)</td>
<td>2.8555</td>
<td>(1.974, 3.7099)</td>
</tr>
<tr>
<td>α4</td>
<td>7.668</td>
<td>(6.0710, 9.2656)</td>
<td>4.6793</td>
<td>(3.4184, 6.0124)</td>
</tr>
<tr>
<td>α5</td>
<td>9.687</td>
<td>(7.9118, 11.4630)</td>
<td>6.0494</td>
<td>(4.4793, 7.6323)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prediction Error</th>
<th>Random intercept model Percentage</th>
<th>Areal model Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>28%</td>
<td>25%</td>
</tr>
<tr>
<td>-1, 0, 1</td>
<td>79%</td>
<td>73%</td>
</tr>
<tr>
<td>-2, -1, 0, 1, 2</td>
<td>96%</td>
<td>94%</td>
</tr>
</tbody>
</table>