Clinical Decision Rules for Detection of Cartilage Degradation Based on Univariate MR Parameter Analysis

Richard G. Spencer[1], Vanessa A. Lukas[1], Benjamin D. Cortese[2], David A. Reiter[1], Kenneth W. Fishbein[1], Nancy Pleshko[3], and Bimal Sinha[4]

[1]*Magnetic Resonance Imaging and Spectroscopy Section, National Institute on Aging, National Institutes of Health, Baltimore, Maryland, United States,* [2]*Department of Mathematics, Syracuse University, Syracuse, New York, United States,* [3]*Tissue Imaging and Spectroscopy Laboratory, Bioengineering Department, Temple University, Philadelphia, Pennsylvania, United States,* [4]*Department of Mathematics and Statistics, University of Maryland, Baltimore County, Baltimore, Maryland, United States*

Introduction: Little work has been done to translate cartilage matrix-sensitive MR outcome measures to clinically relevant decision rules in spite of the fact that MR would be an ideal noninvasive modality to define cartilage status for preclinical studies, for clinical prognosis and staging, and to evaluate matrix-altering therapeutic interventions[1,2]**.** Intuitively, larger differences in parameter values between two groups and smaller scatter within each group will lead to more reliable outcome assignments for unknown subjects. ***The goal of this work is to formalize this concept, and to develop and apply clinical outcome rules based on group differences between cartilage matrix-sensitive MR measurements.***

Theory: We describe two distinct methods for assigning a new subject (S_{new}) with a parameter measurement (p_{new}) to a control (Ctl) or disease (Dis) group based on known estimates of group means, μ_{Ctl} and μ_{Dis}, and standard deviations, σ_{Ctl} and σ_{Dis}. **Empirical Assignment Rules: Euclidean distance metric (EDM):** S_{new} is assigned to the group whose mean value is closer to p_{new}. That is, $|p_{new}-\mu_{Ctl}|<|p_{new}-\mu_{Dis}| \Rightarrow S_{new}\in Ctl$ and $|p_{new}-\mu_{Dis}|<|p_{new}-\mu_{Ctl}| \Rightarrow S_{new}\in Dis$. The EDM does not take into account σ_{Ctl} and σ_{Dis}, which describe how likely it is for a measurement to have a value distant from its group mean. **Likelihood ratio (LR):** The LR incorporates σ_{Ctl} and σ_{Dis}. Using Bayes theorem and assuming, for convenience, equal pre-test group assignment probabilities and Gaussian-distributed data, we find that $\frac{1}{\sqrt{2\pi}\sigma_{Ctl}}e^{\frac{-(p_{new}-\mu_{Ctl})^2}{2\sigma_{Ctl}^2}} > \frac{1}{\sqrt{2\pi}\sigma_{Dis}}e^{\frac{-(p_{new}-\mu_{Dis})^2}{2\sigma_{Dis}^2}} \Rightarrow S_{new}\in Ctl$ and $\frac{1}{\sqrt{2\pi}\sigma_{Ctl}}e^{\frac{-(p_{new}-\mu_{Ctl})^2}{2\sigma_{Ctl}^2}} < \frac{1}{\sqrt{2\pi}\sigma_{Dis}}e^{\frac{-(p_{new}-\mu_{Dis})^2}{2\sigma_{Dis}^2}} \Rightarrow S_{new}\in Dis$. The quality of a binary assignment rule is defined by its sensitivity (SE) and specificity (SP): $SE=\Pr\{p_{new}\rightarrow Dis \,|\, S_{new}\in Dis\}$ and $SP=\Pr\{p_{new}\rightarrow Ctl \,|\, S_{new}\in Ctl\}$, the respective probabilities (Pr) of correctly assigning S_{new} to the Dis or Ctl group based on p_{new}. **Theoretical Assignment Rules:** Closed form expressions for SE and SP of the EDM and LR assignment methods for Gaussian-distributed data can be derived by appropriate integrations over the group probability distributions. Defining $\Phi[v]=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{v}e^{-x^2/2}dx$ one obtains: **EDM:** $SE=\Phi\left[\frac{|\mu_{Ctl}-\mu_{Dis}|}{2\sigma_{Dis}}\right]$ and $SP=\Phi\left[\frac{|\mu_{Ctl}-\mu_{Dis}|}{2\sigma_{Ctl}}\right]$. **LR:** With (C_1',C_2') defined as the roots of $\left(\frac{p_{new}-\mu_{Dis}}{\sigma_{Dis}}\right)^2-\left(\frac{p_{new}-\mu_{Ctl}}{\sigma_{Ctl}}\right)^2-\log\left(\frac{\sigma_{Ctl}}{\sigma_{Dis}}\right)=0$, $C'_{\min(\max)}=\min(\max)(C_1',C_2')$, and $K=sign(\sigma_{Ctl}-\sigma_{Dis})$, one obtains: $SE=\Phi\left[K\frac{C'_{\max}-\mu_{Dis}}{\sigma_{Dis}}\right]-K\Phi\left[\frac{C'_{\min}-\mu_{Dis}}{\sigma_{Dis}}\right]$ and $SP=\Phi\left[-K\frac{C'_{\max}-\mu_{Ctl}}{\sigma_{Ctl}}\right]+K\Phi\left[\frac{C'_{\min}-\mu_{Ctl}}{\sigma_{Ctl}}\right]$.

Experimental methods: *Sample Preparation:* BNC disks (8 mm diameter) threaded onto hollow polyethylene tubes were inserted into a susceptibility-matched ULTEM sample holder bathed in Fluorinert FC-77 and imaged at 25 °C and 3T before and after moderate pathomimetic degradation with trypsin. *MRI Measurements:* T_1, T_2, T_2*, magnetization transfer ratio (MTR), and ADC values were acquired on a 3T Philips Achieva MR system according to our established protocols before and after enzymatic degradation. *Dataset Construction:* To reflect the heterogeneity of the osteoarthritic process, Gaussian-distributed datasets were constructed from multiple trypsin degradation times. *Empirical assessment of SE and SP:* A cross-validation procedure using 100 random 2/3 - 1/3 splits between training and validation data was used to calculate SE and SP according to the empirical assignment rules specified above. These results were compared with the corresponding theoretical assignment rules.

Results: Parameter means (Table 1) and distribution functions (Fig. 1) are shown below. Table 2 shows results for EDM and LR classification of experimental data. In many cases, the significant differences resulted in modest values for SE and SP, of ~ 0.85 or below. General agreement is seen between the experimental and theoretical results, with differences due to the finite and imperfectly Gaussian data. EDM and LR differed most for MTR, which had the largest group differences in σ. Results based on literature values are shown in Table 3. Note again the modest SE and SP for discrimination, in spite of statistically significant group differences.

	T_1 (ms)	T_2 (ms)	T_2* (ms)	ADC (x 10^{-4} mm^2/s)	MTR ($1-M_{ss}/M_0$)
Control (n=24)	811.2 ± 26.3	87.4 ± 8.1	22 ± 4	1.20 ± .08	0.376 ± 0.007
Trypsin (n =150)	910.4 ± 60.2***	127.0 ± 16.7***	30 ± 6***	1.51 ± .11***	0.345 ± 0.027***

*Table 1. Parameter means and standard deviations for experimental dataset. (*** p<.001)*

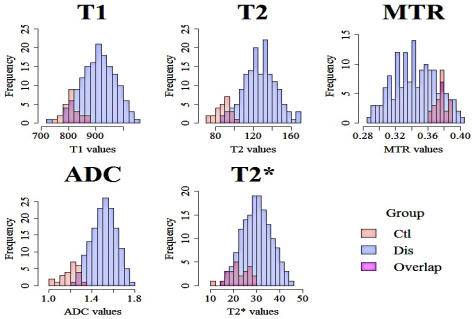

Figure 1. Histograms of the parameters values for control (Ctl) and degraded (Dis) cartilage samples.

		Empirical SE	Empirical SP	Theoretical SE	Theoretical SP
T_1 (ms)	EDM	0.78	0.90	0.80	0.97
	LR	0.80	0.91	0.83	0.94
T_2 (ms)	EDM	0.86	1.00	0.88	0.99
	LR	0.92	1.00	0.93	0.97
T_2* (ms)	EDM	0.73	0.76	0.75	0.84
	LR	0.71	0.82	0.73	0.86
ADC (x 10^{-4} mm^2/s)	EDM	0.89	1.00	0.92	0.97
	LR	0.93	1.00	0.94	0.96
MTR ($1-M_{ss}/M_0$)	EDM	0.69	0.97	0.72	0.99
	LR	0.76	0.92	0.79	0.95

Table 2. Comparison of empirical and theoretical SE and SP of the two clinical tests, EDM and LR.

Study	MR Parameter	Ctl	Dis	SE	SP
Mosher TJ, et al.[3]	T_2 (ms)	45.59 ± 1.92	47.38 ± 2.72*	0.53	0.79
Friedrich KM, et al.[4]	T_2 (ms)	45.7 ± 3.56	50.1 ± 4.52*	0.64	0.78
Yao W, et al.[5]	T_2 (ms)	37.8 ± 3.3	44.0 ± 8.5*	0.64	0.90

*Table 3. Significant group differences (*p<.05) from the literature in terms of SE and SP for LR. All results are for human cartilage data at 3T.*

Discussion: We have shown both empirically and theoretically that a small p value for group differences may not translate into accurate classification. Ultimately, this is because statistical significance can derive from even small differences when enough individuals are sampled (large enough n), while a decision rule for a given patient is ultimately an n = 1 measurement. The formalism developed here provides a means to examine this. The EDM is highly intuitive and leads to a well-defined cutoff value for classification. However, if the σ's of the two groups are very different, it will lead to erroneous results. The LR is the most accurate classifier, but is more complex and results in disjoint decision regions. Our results can be extended non-Gaussian data distributions, and to incorporate random measurement error. We propose that in translational or clinical research, outcomes be accompanied by an assessment of the SE and SP of the results for detection of the binary variable of interest. We believe that in many cases this will serve as a valuable counterpoint to the optimism engendered by the finding of a statistically significant difference between groups.

References: 1) Laupacis AN, et al. JAMA 1997; 277(6):488-494. 2) Spencer RG and Pleshko N J Am Acad Orthop Surg 2013; 21(7):438-439. 3) Mosher TJ, et al. Radiology 2011; 258(3):832-842. 4) Friedrich KM, et al. AJR Am J Roentgenol 2009; 193(5):W411-415. 5) Yao W, et al. Skeletal Radiol 2009; 38(11):1055-1062.