Are two samples of parametric maps statistically different? Indexed distribution analysis (IDA) can provide better inferences than conventional and histogram analysis methods

Chris J Rose[1], James P O'Connor[1,2], Tim F Cootes[1], Chris J Taylor[1], Gordon C Jayson[3], Geoff J Parker[1], and John C Waterton[1,4]

[1]*Center for Imaging Sciences, Manchester Academic Health Science Center, The University of Manchester, Manchester, Greater Manchester, United Kingdom,* [2]*Department of Radiology, The Christie, Manchester, Greater Manchester, United Kingdom,* [3]*Department of Medical Oncology, The Christie, Manchester, Greater Manchester, United Kingdom,* [4]*AstraZeneca, Alderley Park, Cheshire, United Kingdom*

TARGET AUDIENCE Researchers who use MR biomarkers (e.g., T_1, T_2, K^{trans}, ADC, etc.) in natural history studies, or clinical or pre-clinical trials of novel therapies.

PURPOSE To validate the recently proposed indexed distribution analysis (IDA) method[1] for analyzing spatially heterogeneous samples of parametric maps using well-controlled simulated and clinical imaging data, and to compare IDA to conventional and histogram analyses.

METHODS If spatial correspondences cannot be established across a sample (e.g. using an atlas), parametric maps of an imaging biomarker are often reduced to a scalar summary (e.g. an average) to test hypotheses (e.g. of no difference in K^{trans} between pre- and post-treatment conditions). However, such summaries might fail to reflect the underlying biology, and may be insensitive to population differences or treatment effect. Common alternatives include various histogram analysis methods, for example in which the heights of corresponding histogram bins are subjected to significance testing[2].

IDA establishes approximate correspondences between features of parameter distributions across each map in a sample. Hypotheses of no difference in corresponding parameter values can then be tested to infer the magnitude, direction, and significance of differences. Resulting *P*-values can be mapped into the space of the original images to identify tissues that likely differ between experimental conditions. In paired experiments, a confidence interval (CI) on the proportion of voxels that differ between conditions estimates the mean *spatial extent* of differences (e.g. treatment-induced effect).

Two experiments, using data from a simulated study and from a clinical study of bevacizumab[3], were performed. In each, two pre-treatment scans (Pre1 & Pre2) provided negative control and a post-treatment scan (Post) provided positive control. Three analyses were performed for each experiment: conventional significance testing of parameter averages using paired two-sided *t*-tests; histogram analysis using paired two-sided *t*-tests; and IDA. False discovery rate (FDR) control[4] was used in histogram analysis and IDA to minimize the risk of type I errors.

Fig. 1 (top three rows) shows data for the simulated study. Structures were comprised of two simulated tissues with parameter values sampled from $\mathcal{N}(1, 1/4)$ and $\mathcal{N}(10, 1/4)$ respectively at Pre1. For each structure, inter-structure shift in parameter value was modeled by adding a single sample from Gamma(2, 1/2) to each pixel. Inter-scan measurement error was simulated at Pre2 by sampling from $\mathcal{N}(0, 1/4)$. For each structure, treatment-induced change was simulated in the central tissue by subtracting a sample from $\mathcal{N}(4, 1/4)$ from the Pre2 maps. For conventional significance testing, structures were summarized by means; 14 histogram cut-points were used.

The clinical experiment used retrospective K^{trans} data (unit: min^{-1}) from a study of bevacizumab in 10 patients with a total of 26 colorectal carcinoma liver metastases. Data for one patient (3 tumors) were omitted, as data for the patient's first scan was unavailable. Pre1 and Pre2 data were acquired in the week preceding dosing; Post data were acquired 48 hours after dosing. Voxels satisfying $0 \le K^{trans} \le 1.5$ were included. As the K^{trans} data were skewed, analyses were performed on log K^{trans} (base e). For conventional significance testing, structures were summarized by medians; 12 histogram cut-points were used. Research ethics committee approval was granted; written informed consent was obtained.

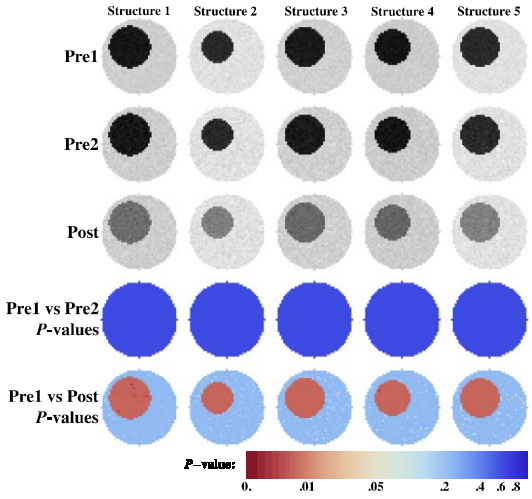

Fig. 1 *Parameter and P-value maps for the simulated imaging study (see text).*

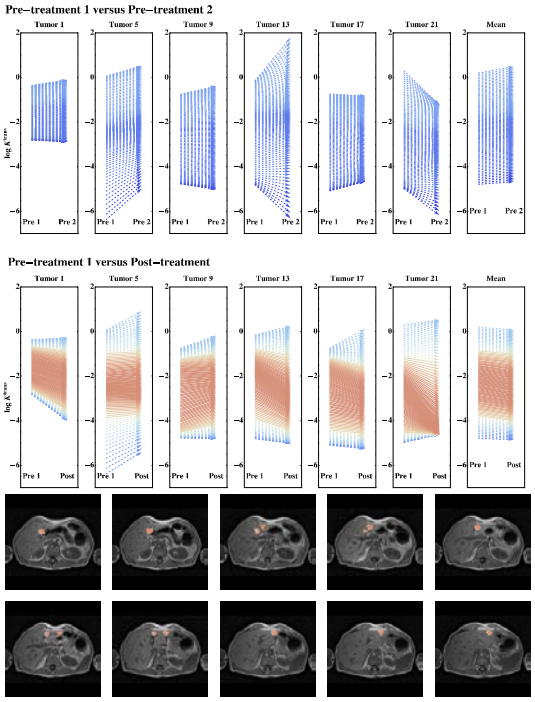

Fig. 2 *Slopegraphs for the bevacizumab study and P-value maps for three tumors in one patient on a T_1-weighted Pre1 image. See Fig.1 for P-value color map.*

Table 1 *P-values and 95% confidence intervals on the estimated quantities (see text) for the two experiments.*

		Conventional analysis	Histogram analysis	Indexed distribution analysis
Simulated study	**Pre1 vs. Pre2**	$P = .808$ (-.00505, .00609)	$P > .05$ for all bins	P = Undefined (see text)
	Pre1 vs. Post	$P = 5.50\times10^{-4}$ (.743, 1.31)	$P > .05$ for all bins	$P = .00101$ (.187, .320)
Bevacizumab study	**Pre1 vs. Pre2**	$P = .114$ (-.172, .0197)	$P > .05$ for all bins	$P = 9.4\times10^{-20}$ (5.2×10^{-4}, 1.3×10^{-3})
	Pre1 vs. Post	$P = .00991$ (.0608, .398)	$P > .05$ for all bins	$P = 2.1\times10^{-19}$ (.943, .962)

RESULTS Table 1 presents *P*-values and 95% CIs on the mean difference in average imaging parameter value (conventional analyses), *P*-value summaries (histogram analyses), and *P*-values and 95% CIs on the spatial extent of differences (IDA). Histogram analysis failed to detect pronounced simulated and known treatment effects. Conventional analysis and IDA *correctly did not* detect differences between Pre1 and Pre2, and *correctly did* detect differences between Pre1 and Post. The IDA result for the Pre1 and Pre2 comparison in the simulation study is undefined because no pixels were inferred to have changed leading to a division by zero (Fig. 1, row 4; uniform *P*-values are an artifact of FDR control). However, the correct inference to draw is that the mean spatial extent of simulated treatment is zero. The conventional analysis dramatically under-estimated simulated treatment effect magnitude (as .743–1.31; Table 1). IDA identified pixels affected by simulated treatment with 100% accuracy (Fig. 1, bottom row) and correctly inferred treatment effect magnitude in these pixels to be approximately 4 (consistent with the mean value of the distribution used in the simulation). IDA appears to infer a difference between Pre1 and Pre2 in the bevacizumab study (Table 1). However, the 95% CI on the mean spatial extent of bevacizumab-induced change is essentially zero. IDA inferred that K^{trans} (c.f. log K^{trans}) decreases by 25% after treatment (Fig. 2, mean slopegraph), consistent with the literature[3]. IDA inferred that bevacizumab affects 94–96% of the tumor volume (Table 1), consistent with current understanding.

CONCLUSIONS Histogram analysis can be insensitive to pronounced treatment effects. Conventional average-based analyses can dramatically underestimate treatment effect magnitude. IDA can correctly estimate treatment magnitude and spatial extent, and allows inferences to be visualized with respect to each structure.

REFERENCES **1** Rose C. et al., Proc. ISMRM #755, 2012. **2** de Lussanet et al., Int J Radiat Oncol 2005;63(5):1309–15. **3** O'Connor et al., Clin Cancer Res 2009;15(21):6674–82. **4** Benjamini et al., Ann Statist 2001;29(4):1165–88.

ACKNOWLEDGEMENTS This work was funded by AstraZeneca and The University of Manchester.