Are two samples of parametric images statistically different? Novel significane tests on samples of density estimates, with application to ventilation-to-perfusion mapping of the lung in COPD using oxygen-enhanced MRI

Chris J Rose[1,2], Penny L Hubbard[1,2], Tim F Cootes[1,2], Chris J Taylor[1,2], Josephine H Naish[1,2], Geoff J Parker[1,2], Simon S Young[3], and John C Waterton[1,4]
[1]*University of Manchester Biomedical Imaging Institute, Manchester, Greater Manchester, United Kingdom,* [2]*University of Manchester Academic Health Science Centre, Manchester, Greater Manchester, United Kingdom,* [3]*AstraZeneca R&D, Charnwood, Loughborough, Leicestershire, United Kingdom,* [4]*AstraZeneca, Alderley Park, Macclesfield, Cheshire, United Kingdom*

INTRODUCTION MRI biomarkers are used in natural history and intervention studies to spatially map physical or physiological parameters throughout a structure. Given a parametric map for each individual in a sample, the problem of choosing the optimal statistical method for drawing inferences about the populations being compared (pre- vs post-treatment; healthy vs diseased) is unsolved. For each individual, multiple measurements are made (e.g., over thousands of voxels) on structures that may not be homogeneous in the parameter studied. The common approach of summarizing each map by its average, and using those in a hypothesis test can be flawed: averages are insensitive to changes in small but possibly important regions, and may not reflect underlying biology (consider an average of data drawn from a distribution with two equally-sized modes). Sometimes, the problem can be addressed by establishing spatial correspondences across the sample, facilitating comparison at corresponding locations (e.g., the use of atlases in neuroscience). But this is intractable if the structure of interest varies excessively within or between individuals. An example is ventilation-to-perfusion (*V*/*Q*) mapping of the lung using O_2-enhanced MRI[1]. Histogram analysis[2] appears promising, but is practically difficult. Bin heights or locations cannot be compared across the sample: distributions can translate and scale relative to one another; modes in distributions can translate non-systematically across the sample; modes may appear, grow, shrink, disappear or merge due to intervention; and comparing many bins engenders multiple comparison issues that, if unaddressed, will result in high type I error rate. We have developed significance tests that can be applied to a sample of density estimates in this setting.

THEORY Assume a sample of parametric maps obtained under conditions A and B (e.g., pre- & post-treatment). The i^{th} map, X_i, contains n_i univariate measurements (one per voxel) of an MR parameter, x (e.g., *V*/*Q*, T_1, etc.). We seek to establish approximate correspondences between values of x across the sample by fitting a reference distribution, $D_{\text{ref}}(t)$, $0 \le t \le 1$, to data from X_i, which we assume was drawn from distribution $D_i(x)$. Having established the correspondences, hypothesis tests can be applied: for example, we may test the hypothesis that the mean volume of tissue giving rise to a particular feature (e.g., a common mode) is equal under conditions A and B. (In practice we do not identify or name features, but apply such tests across a range of corresponding values of x). Correspondences are established by nonlinearly transforming values of t—in the domain of D_{ref}—to "warp" D_{ref} to match distribution D_i, using transform $\omega_{\boldsymbol{\theta}_i}$. The values of x and D_i corresponding to a value of t, for the i^{th} map, can be written parametrically as:

$$x = \omega_{\boldsymbol{\theta}_i}(t) \quad \& \quad D_i = D_{\text{ref}}(t)/\omega_{\boldsymbol{\theta}_i}{}'(t). \qquad (1)$$

Dividing by derivative $\omega_{\boldsymbol{\theta}_i}{}'$ ensures that the warped version of D_{ref} integrates to unity (i.e., is a legal distribution). Parameter vector $\boldsymbol{\theta}_i$ includes coefficients that allow D_{ref} to shift and scale as a whole, and coefficients that describe the nonlinear aspect of the transform. The nonlinear aspect of $\omega_{\boldsymbol{\theta}_i}$ is based on weighted sums of integrals of Bernstein polynomials[3]; the integrals can be pre-computed (a computationally expensive task), allowing transforms to be applied many times during the estimation of $\boldsymbol{\theta}_i$ using a (computationally cheap) matrix multiplication. Each $\boldsymbol{\theta}_i$ is estimated in a Bayesian manner using Markov Chain Monte Carlo, but alternatives like maximum likelihood may also be used. Corresponding values of x in each D_i can then be found by computing $x = \omega_{\boldsymbol{\theta}_i}(t)$ for a given value of t (i.e., a given value of t "means the same thing" for all maps in the sample).

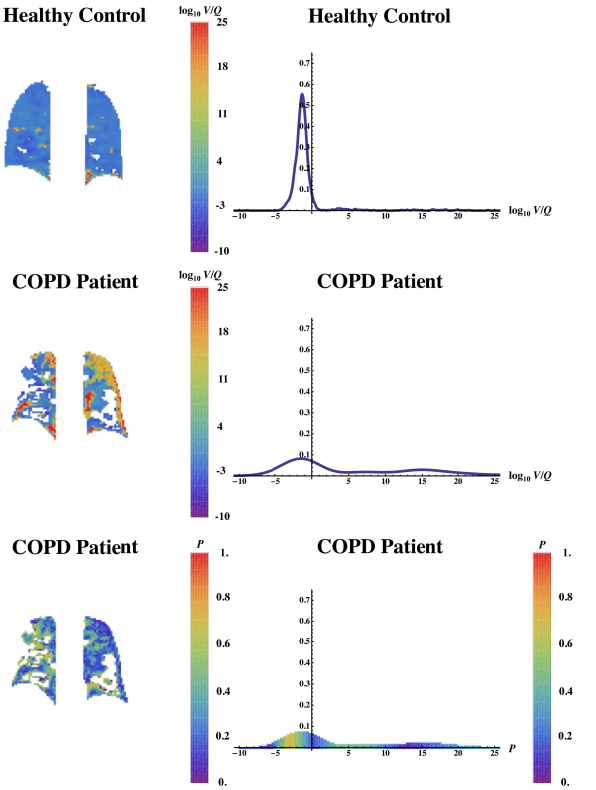

Fig. 1 *Parametric maps and kernel density estimates of ventilation to perfusion ratio (*log *V/Q) in a healthy control (top row) and a patient with severe COPD (middle row). While a conventional t-test is insensitive to the heterogeneous differences between the two populations, the proposed method can do. By establishing correspondences between density estimates across the study, we observe in the COPD patients a significant decrease in the volume of tissue with normal V/Q and a significant increase in the volume of tissue with abnormally high V/Q. The bottom row shows P-values for the same COPD patient (see text for full details).*

If A and B represent pre- and post-intervention conditions (i.e., a paired experiment), we can test hypotheses that there are no differences in the locations of corresponding features (i.e., values of x). If A and B represent groups in a parallel study (unpaired), we can test the hypotheses that there are no differences in the volumes of tissues underpinning corresponding features. This is achieved by calculating the volumes of tissue with values of x between corresponding lower and upper bounds, sequentially across the sample, and then applying t-tests. The multiple comparisons issue can be addressed by controlling the false discovery rate[4].

METHOD In a study of chronic-obstructive pulmonary disease (COPD), O_2-enhanced MRI (1.5T Philips Achieva, Philips Medical Systems, Best, NL) and tracer kinetic modeling were used to spatially map ventilation-to-perfusion ratio (*V*/*Q*, usually expressed on a log scale) in 12 patients with severe COPD (Gold stage 3–4) and 11 healthy age-matched controls. Written informed consent was obtained. We compared the patients and controls in two ways: conventionally, by summarizing each map by its median and applying a Mann-Whitney *U*-test; and as described above, testing hypotheses of no differences in the volumes of tissues underpinning corresponding features.

RESULTS Fig. 1 shows maps of log *V*/*Q*, and corresponding kernel density estimates, for a healthy volunteer and a COPD patient. Although the parameter maps and distributions of log *V*/*Q* are visually distinguishable in most cases, the *U*-test was non-significant ($P = 0.713$). Fig. 1 shows a map of the *P*-values calculated by the method described above for the same patient, and those *P*-values mapped onto a kernel density estimate. The volume of healthy lung tissue (log *V*/*Q* $\approx$ 0) was statistically significantly lower in the COPD patients, and the volume of tissue with abnormally high log *V*/*Q* (values 10–15 in patient 1531) was statistically significantly higher in the COPD patients; on average, each lung had 10 voxels whose associated *P*-values were strictly less than 0.05 (i.e., statistically significant differences were identified in an average of 1220.5 mm^3 of lung volume). The resulting *P*-value for each corresponding "feature" can be displayed on the density estimates, and can be mapped onto the original images to show the spatial locations of tissues that are statistically different between healthy volunteers and COPD patients (see bottom row in Fig. 1).

CONCLUSIONS We have developed a method for analyzing heterogeneous changes or differences arising in imaging studies and applied it to MR data from a study of COPD. Unlike the conventional method of summarizing parameter maps using an average and using those in a hypothesis test, our method can identify biologically meaningful statistically significant differences in ventilation-to-perfusion ratio between severe COPD patients and age-matched healthy controls. Further, our method does not require the identification of "hot spots" or other investigator input, and so is completely objective. While we have applied the method to *V*/*Q* data from an O_2-enhanced MRI study, it is generically applicable to many other MR imaging parameters, including dynamic contrast-enhanced and diffusion MRI.

REFERENCES **1** Naish 2010, Proc. ISMRM, #2516. **2** Yankeelov 2007, Magn Reson Imag 25(1):1–13. **3** Davies 2008, Statistical Models of Shape, Springer. **4** Benjamini 2001, Ann Statist 29(4):1165–1188.

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