1

Using NMR Images as the Primary Outcome in Clinical Trials

Nick Fieller

Department of Probability & Statistics
University of Sheffield, UK

2

- **Joint work with Emma O'Connor**

- **Acknowledgements**

John Waterton
Andrew Holmes
& members of the Discovery, Oncology & Imaging Depts.
AstraZeneca

3

Outline

- Introduction
- Exploratory analysis
  - ♦ histograms & kernel densities
- Initial analysis
  - ♦ sample statistics, medians, K-S tests
- Main analysis
  - ♦ functional principal components & randomization tests
- Loose ends & refinements

4

Introduction

- 3 experiments:-
  change in vascular permeability in tumours after drug designed to inhibit angiogenesis
- Four doses
  - ♦ 12.5mg, 25mg, 50mg, 100mg + control
- 3D images of tumours by MRI before & 24 hrs after treatment
- Data provided:-

voxel values of K^{trans}

5

Generic problem:

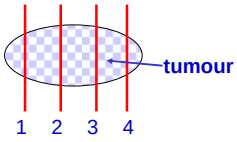
- ♦ Measured response is an **'image'**
  - i.e values of 1 (or more) parameter(s) in pixels/voxels (+ spatial coordinates)
- ♦ General statistical objective:– explore avenues for statistical analysis of experiments when responses are images
- ♦ **N.B.** Not analyzing images *per se*

6

- **Key Advantages:–**
  - ♦ NMR Images are non-invasive
  - ♦ Subject can be monitored through time
    - Other measures of tumours require sacrifice of subject
  - ♦ Clear statistical advantages
    - Lessens ***inter-subject*** variability
    - Treatment comparison is ***intra-subject***
  - ♦ + ethical advantages
  - ♦ May demonstrate clinical effects quickly
    - c.f. other clinical outcomes, e.g. 5-year survival

7

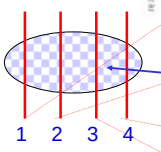
'Raw Data'

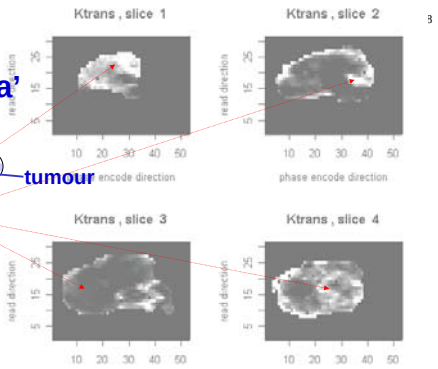


NMR scan focuses separately on four near-adjacent 'slices' measuring parameters in each voxel

8

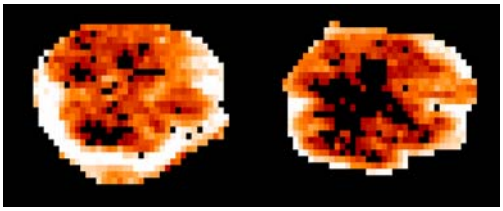
'Raw Data'





K^{trans} Maps of all four slices of *E12M48pre*

9



Maps of K^{trans} before and after treatment

E14M34: 50mg dose

Images of tumour before and after treatment

Lighter colour indicates higher permeability :–

note visible decrease in permeability post treatment

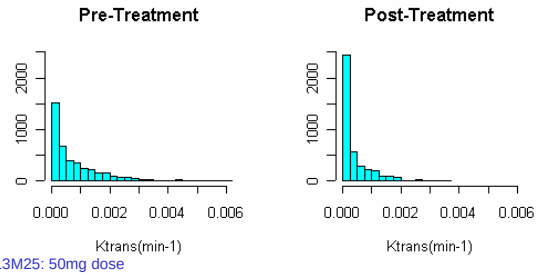
10

Exploratory analysis

- 3D images provided as four 2D slices
  - Decide to use only 'middle 2 slices'
    - Outer slices 'almost empty' in many images
  - Many zero values of K^{trans}
    - (internal to image – ROI excludes 'external zeros')
  - Histograms of voxel values have very high positive skewness
  - $\Rightarrow$ work with log(K^{trans})
    - or log(1+ K^{trans}) or **log(ε+ K^{trans})** to avoid log(0)

11

- **Notes**
  - log(1+ K^{trans}) or log(ε+ K^{trans}) ?
    - DOES NOT REALLY MATTER
    - (ε < half resolution of measurement)
  - Lack of registration between images
    - images 'before' & 'after' treatment are not aligned or 'registered' so cannot look at changes at voxel level
      - tumours are non-rigid
  - Lack of registration between slices
    - 2D slices are not perfectly registered so at best only partial information on proximity of voxels in different slices



E13M25: 50mg dose

Histograms of pixel values in slices 2 & 3 pooled

(note overall decrease in values post-treatment)

13

- **Notes:**
  - Measured response:– Image
  - Working unit:– sample distribution of K^{trans}
  - Losses
    - spatial information
      - (but see later loose end)
  - Gains
    - more familiar statistical object
    - allows calculation of 'mean responses' & assess variability relative to mean
      - (by pooling sample distributions)

14

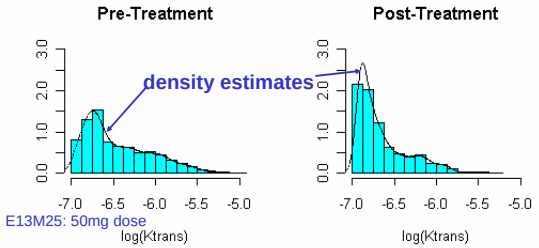
- **Notes(ct$^{d.}$)**
  - Ignoring spatial information masks ***dependence*** of voxel values
    - K^{trans}–values are not independent
    - Beware of statistical techniques which assume independence
    - 'sample size' is ***NOT*** #voxels
    - true sample size is #images, i.e. 88 in total from 44 individuals ~ 6 in each dose group

15

- **Notes(ct$^{d.}$)**
  - Working with distribution looses information on ***size*** of tumour
    - size change in 24 hours implausible
      - but could be an issue over weeks/months
  - Beware of effects of zero values of K^{trans}
    - (trace zeroes ←→ structural zeroes)
    - reliant on accuracy of segmentation techniques to determine ROI

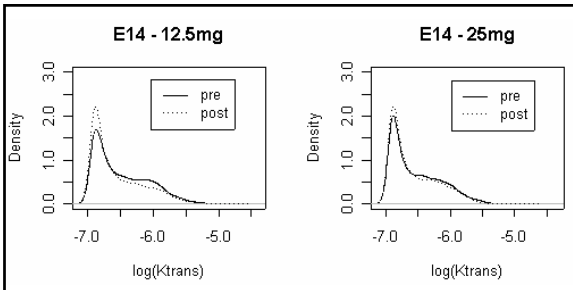
16

- *kernel density estimates* better than histograms
  - gives a 'smooth' version of the histogram



17

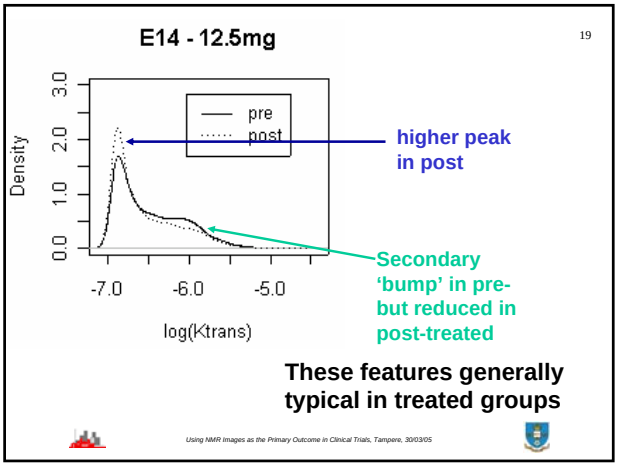
- **Notes:**
  - measured response: **image**
    - ⇓
    - sample distribution
    - ⇓
    - histogram
    - ⇓
    - **Working unit:– kernel density estimate**
  - **Gains:–** reduced random noise & comparability between samples
    - use common smoothing parameter
  - **Losses:–** actual values



Average densities for 12.5mg and 25mg dose groups

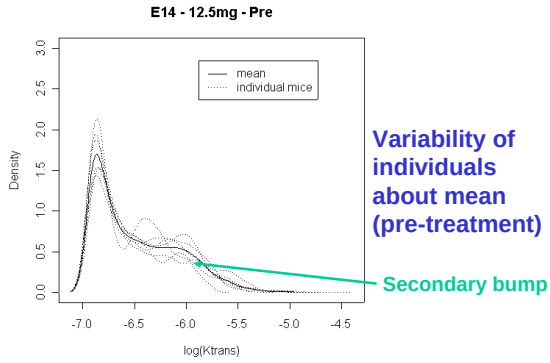
Note visible overall decrease in values post-treatment

19

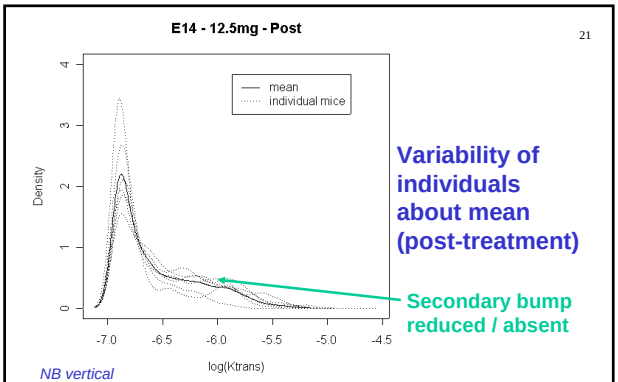


These features generally typical in treated groups

20



21



22

Initial Analyses

- Sample medians
  - Generally lower post-treatment in all four treated dose groups
    - some anomalies (6 of the 33 treated subjects)
  - No pattern in control groups
    - 4 decreased, 7 increased

23

- Kolmogorov-Smirnov 'tests'
  - tests for differences in sample distributions
  - very powerful test with large sample sizes
  - most tests of pre- *vs* post-treatment indicated statistically 'significant' differences (***except*** in control groups)
    - NB test presumes *independent observations* which is clearly not true in this case
      - p-values ***artificially*** small
    - absence of difference in control groups is of note

24

- **Comments on Initial Analyses**
  - simple sample summary statistics only capture changes in **location** and **scale**
  - exploratory analyses (kernel densities) shewed more complex changes
  - Kolmogorov-Smirnoff tests reveal small differences in distribution but say nothing about type of difference
    - not possible to see whether difference is clinically interesting or which sample is 'preferable'

25

Main Analyses

- Starting point:
  - ♦ measured response is a probability density
    - image ⇒ distribution ⇒ histogram ⇒ **density**
  - ♦ probability densities are **functions**
  - ♦ Key references:

Jim Ramsay & Bernard Silverman:
Functional Data Analysis
Applied Functional Data Analysis
(Springer, 1997 & 2002)

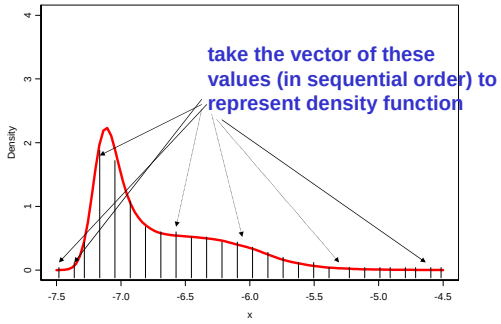
26

- **Key ideas:**
  - ♦ investigate components of **variability** between the functions
    - use a type of principal component analysis
  - ♦ relate these components to differences such as pre- / post- treatment or control / treated subjects or different dose levels

27

- **First step:**
  - ♦ need to summarize / parameterize the functions in a small number of values or parameters
    - note that kernel density estimates are *non-parametric* estimates
  - ♦ R & S recommend using a basis of B-splines to estimate log(density)
  - ♦ We use a simplistic approach and 'discretize' the density by taking its value at 100 equally spaced points

28



29

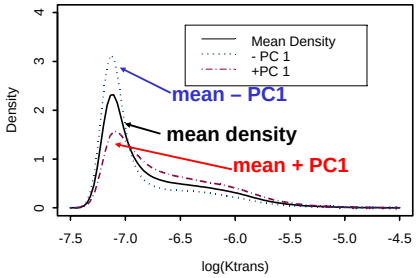
- measured response is a **vector**
  - ♦ image ⇒ distribution ⇒ histogram ⇒ density ⇒ **vector of 100 values**
  - ♦ **Losses:–** details of function by discretization
    - sensitivity analysis using grid of 50 or 200 values shewed little difference in results
  - ♦ **Gains :–** all standard multivariate techniques such as PCA available
  - ♦ **But:–** need to interpret results of PCA in terms of density functions

30

- Functional principal component analysis
  - ♦ PCA on vector of 100 discrete values
  - ♦ Obtain first few (2 or 3) PCs
    - these are also vectors of 100 values
  - ♦ To interpret a PC we need to examine 'typical' densities that are varied from the overall mean in the direction of that PC
    - can do this by constructing a density function as overall mean density +/– PC
    - i.e. add / subtract PC vector from mean vector and then plot values as a density function

PC 1 (77.6%)

31



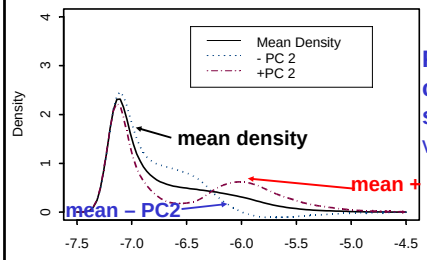
FPCA on complete data set: 77.6% of variation in PC1

PC1 reflects change in height of primary mode

low scores on PC1 ⇔ high peak, high score ⇔ low peak

PC 2 (8.5%)

32



FPCA on complete data set: 8.5% of variation in PC2

PC2 reflects presence of secondary 'bump'

low scores on PC2 ⇔ no bump, high score ⇔ bump

33

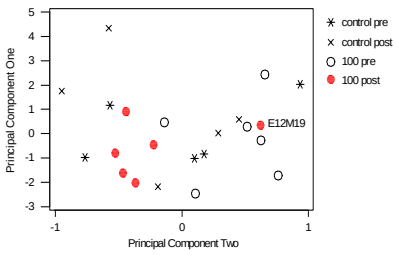
- FPCA (continued)
  - Signs of PCs are arbitrary:–
  - we have chosen signs so low values are **'desirable'** in current context
  - low PC1 ⇔ high peak of low values
  - low PC2 ⇔ no second 'bump' of high values
  - low scores on first 2 PCs ⇔ generally lower permeability & especially fewer fairly high values

34

- FPCA (continued)
  - Can calculate scores for each probability density on the PCs
  - image ⇒ distribution ⇒ histogram ⇒ density function ⇒ vector of 100 values **⇒ scores on first two PCs**
  - **Gains:–** can plot samples in a scatter plot
    - points towards bottom left corner are 'good'
      - (since they have low scores on PCs)
    - Can compare 'pre-' and 'post-' treated

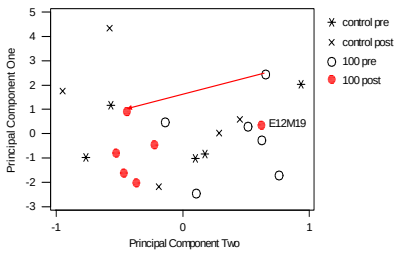
    Discarding low order PCs reduces random noise

35



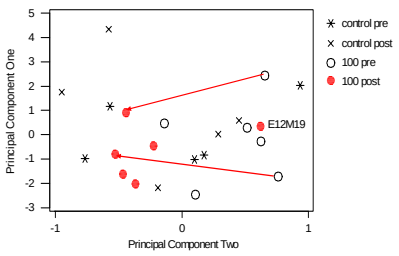
Score plot on first 2 PCs for Expt 12

36



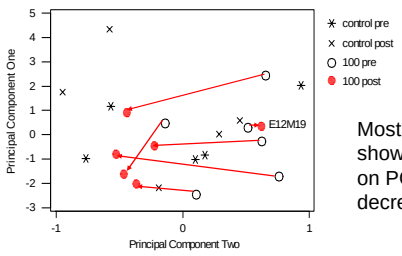
Score plot on first 2 PCs for Expt 12

37



Score plot on first 2 PCs for Expt 12

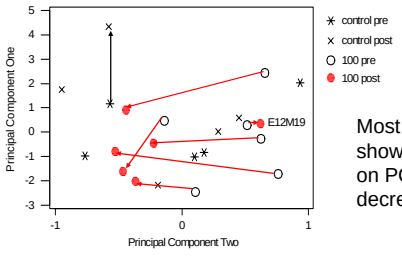
38



Most treated subjects show decrease in score on PC2 and overall decrease in PC1 score

Score plot on first 2 PCs for Expt 12

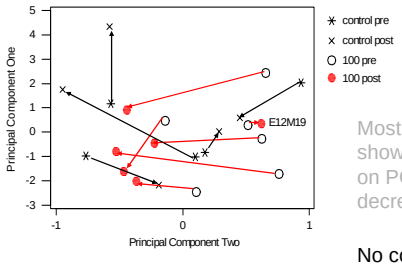
39



Most treated subjects show decrease in score on PC2 and overall decrease in PC1 score

Score plot on first 2 PCs for Expt 12

40

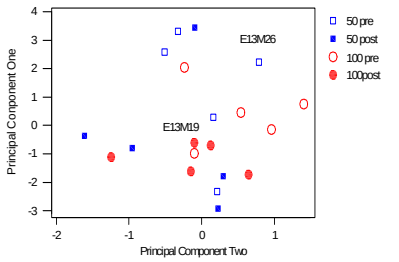


Most treated subjects show decrease in score on PC2 and overall decrease in PC1 score

No consistency in direction of change in controls

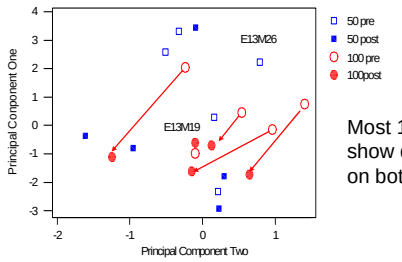
Score plot on first 2 PCs for Expt 12

41



Score plot on first 2 PCs for Expt 13

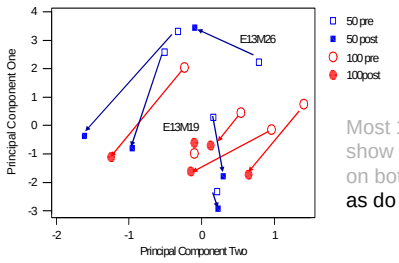
42



Most 100mg subjects show decrease in scores on both PC1 & PC2

Score plot on first 2 PCs for Expt 13

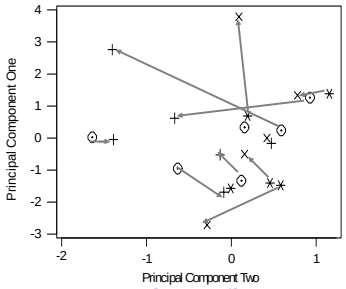
43



Most 100mg subjects show decrease in scores on both PC1 & PC2
as do most 50mg subjects

Score plot on first 2 PCs for Expt 13

44



No consistent pattern of change for control subjects

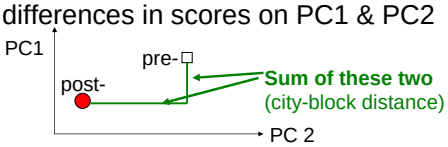
Score plot on first 2 PCs control subjects

45

- **Interim conclusions**
  - overall picture is that treated subjects shew a reduction in permeability both in overall level and in distribution of levels
    - higher proportion of low values
    - lower proportion of fairly high values
  - Control subjects shew no consistent change over 24 hours

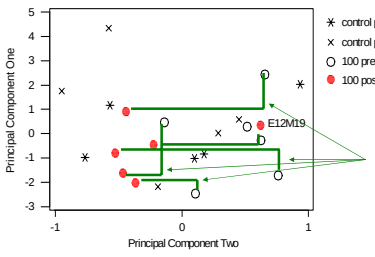
46

- Formal statistical testing:–
  - Calculate test statistic
  - Assess significance by a randomization test based on random relabelling of pre- & post-treated values
- Test statistic based on sum of differences in scores on PC1 & PC2



Sum of these two
(city-block distance)

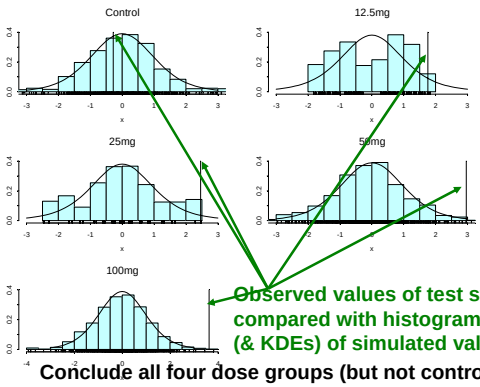
47



test statistic is sum of all these distances

Compare observed value with values from randomly interchanging pre- & post- labels
(near equivalent of paired t-test)

48



Observed values of test statistic compared with histograms (& KDEs) of simulated values

Conclude all four dose groups (but not controls) shew clear evidence of effect of treatment

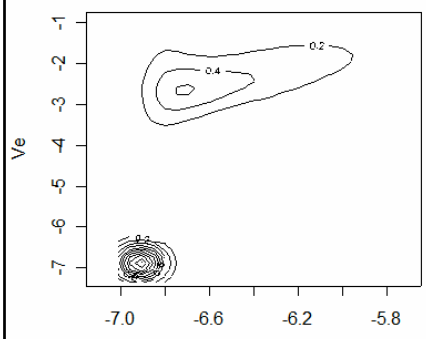
What Next?

9

- Theory extends to two (+?) dimensions
  - Experiment measured 2 parameters in each voxel
    - i.e. bivariate 3D image
    - - Use K^{trans} and V_e simultaneously
    - Need 2d kernel density estimate but otherwise FPCA works as before

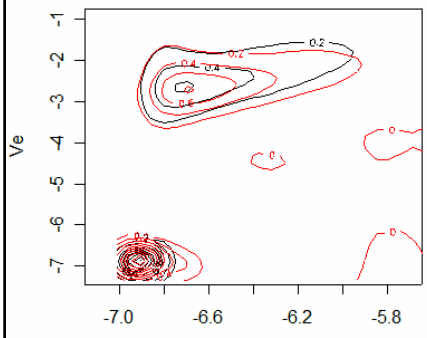
50

E12 - mean



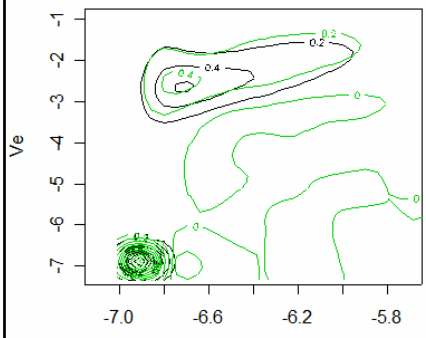
51

E12: mean - PC2



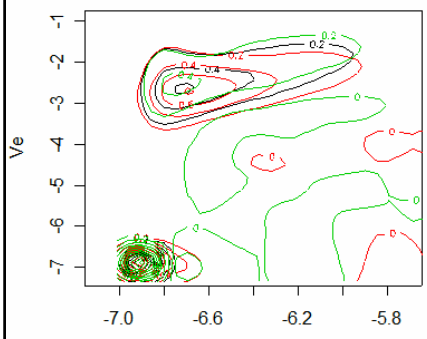
52

E12: mean + PC2



53

E12: mean +/- PC2



PC 2 shows change in V_e